Equilibration of deep neural networks and carrier chirality in Rashba systems

Philipp Christopher Verpoort

Department of Physics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Trinity Hall

December 2020

Copyright Information

Preface

I hereby declare that, except where specific reference is made to the work of others, the content of this thesis is original and has not been submitted in whole or part for consideration for any other degree or qualification at the University of Cambridge or any other university. This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Academic Acknowledgements below and specified in the text. This thesis contains fewer than 60,000 words (including abstract, tables, footnotes and appendices).

The work presented in this thesis is based on the following articles, which have been published in peer-reviewed scientific journals.

[142] Archetypal landscapes for deep neural networks
P.C. Verpoort, A.A. Lee, and D.J. Wales
Proc. National Academy of Sci. USA 117, 21857 (2020).
The results from this work are presented in Chaps. 2 and 3.

[100] Long-lived nonequilibrium superconductivity in a noncentrosymmetric Rashba semiconductor
V. Narayan, P.C. Verpoort, J.R.A. Dann, et al.
Phys. Rev. B 100, 024504 (2019).
The results from this work are presented in Chap. 5.

[143] Chirality relaxation in low-temperature strongly Rashba-coupled systems
P.C. Verpoort and V. Narayan
J. Phys.: Condens. Matter 32, 355704 (2020).
The results from this work are presented in Chap. 6.

As part of my doctoral studies, I have published the following additional articles in peer-reviewed scientific journals, which are not included in this thesis.

[30] Perspective: new insights from loss function landscapes of neural networks
S.R. Chitturi, P.C. Verpoort, A.A. Lee, et al.
Mach. Learn.: Sci. Technol. 1, 023002 (2020).

[34] Au-Ge Alloys for Wide-Range Low-Temperature On-Chip Thermometry
J.R.A. Dann, P.C. Verpoort, J. Ferreira de Oliveira, et al.
Phys. Rev. Applied 12, 034024 (2019).

[141] Materials data validation and imputation with an artificial neural network
P.C. Verpoort, P. MacDonald, and G.J. Conduit
Comp. Mater. Sci. 147, 176 (2018).

[122] Effective-range dependence of two-dimensional Fermi gases
L.M. Schonenberg, P.C. Verpoort, and G.J. Conduit
Phys. Rev. A 96, 023619 (2017).

Academic Acknowledgements

The design of the research and the interpretation of the obtained results in Chap. 3 as well as the review of existing studies in Chaps. 2 and 3 were conducted in collaboration with David J. Wales and Alpha A. Lee.

The computational results presented in Chap. 3 were created in collaboration with and under considerable support by David J. Wales, who ran many GMIN, OPTIM, and PATHSAMPLE jobs and contributed to the creation of graphs, tables, and figures.

The design of the research and the interpretation of the obtained results of the work presented in Chap. 5 was done in collaboration with Gareth J. Conduit, Vijay Narayan, and James R.A. Dann.

The experimental evidence of the magnetoresistance behaviour of GeTe presented in Chap. 5 was obtained by Vijay Narayan and James R.A. Dann.

The research presented in Chap. 6 was conducted by myself with helpful advice from Vijay Narayan.

Philipp Christopher Verpoort
December 2020

Abstract

Title: Equilibration of deep neural networks and carrier chirality in Rashba systems
Author: Philipp C. Verpoort

This thesis reports results of studies conducted on the equilibration of two systems and consists of two parts: the first part deals with the optimisation of deep neural networks, whereas the second part with the decay of non-equilibrium states in strongly Rashba-coupled systems at low temperature.

Deep learning is a conceptually simple, highly effective, and widely used tool, yet there remains insufficient understanding for why it works. The optimisation of deep neural networks with common algorithms such as stochastic gradient descent performs unexpectedly well given the complexity of the underlying high-dimensional non-convex minimisation problem. The first part of this thesis therefore looks at the optimisation procedure from the perspective of statistical physics. This allows us to interpret the loss function landscape of deep neural networks as the counterpart of the potential energy landscape in molecular systems and the optimisation of the network as its equilibration dynamics. Using landscape exploration tools developed in theoretical chemistry, we resolve the structure of the loss function landscape, from which we can draw conclusions for the relaxational dynamics of typical optimisers and, consequently, for deep learning.

The second part investigates how a non-equilibrium charge-carrier chirality distribution in a clean, strongly Rashba-coupled system at low temperatures decays over time. We first motivate this analysis based on experimental studies of transport properties in Rashba materials at low temperatures and subject to external magnetic fields. We investigate whether chirality imbalances could serve as the source for those experimental observations and develop a framework that models the behaviour of such a system. We then proceed with a more general theoretical study of the equilibration mechanisms of chirality in low-temperature strongly Rashba-coupled systems and compute the relaxation timescales of those mechanisms.

Meinen Eltern, die mich immer unterstützt haben.

Acknowledgements

The work reported in this thesis has greatly benefited from invaluable support by many colleagues and friends as well as from an intellectually stimulating research environment in TCM, Trinity Hall, and Cambridge in general. It has been a rewarding experience for me, which I am grateful for having had the privilege to enjoy, and which I will keep in good memory.

I would like to greatly thank those people who have shaped my path during my time at Cambridge, helped me make the difficult decisions I faced, and gave me the guidance and support I needed. First and foremost, I would like to thank Gareth J. Conduit for supervising me during my PhD degree. I have learned much from him, in particular about science, academia, and life in general. I would like to express my sincere gratitude to Vijay Narayan, who has been an enormous help throughout my PhD. His continued support as a colleague and as a friend has been very much appreciated and will be remembered. Moreover, I would like to extend my thanks to David J. Wales for his patience and support throughout our fruitful collaboration. I have learned much from David and have always appreciated his directness and honesty. Equally, I would like to thank Alpha A. Lee for supporting me in completing a successful research project and for sharing his insights into academia, research, and industry, which were useful for determining my future career path. I am grateful to Nick Bampos for helpful advice during my time at Trinity Hall. His advice has been an inspiration and gave me the courage to pursue many projects within and beyond my degree, which have since become great successes. I would also like to thank Jörg Schmalian for encouraging me to apply to Cambridge in the first place, for giving me helpful advice on my studies and my career path, and for being a great inspiration in general.

My research has benefited from many highly capable, friendly, and helping colleagues, including James Dann, John Morgan, and Sathya Chitturi. Michael Rutter saved me whenever an insoluble IT problem emerged, for which I have been very grateful. I would like to thank the TCM group for creating a supportive and enjoyable research environment. My studies and my research would of course not have been possible without the financial support of the Engineering and Physical Sciences Research Council, which is gratefully acknowledged.

Some people made my time at Cambridge particularly enjoyable. I would like to thank Victor Jouffrey for making fun of me whenever I deserved it and for many enlightening conversations, which I will miss dearly. Equally, I am grateful to Bart Andrews for being a friendly and supporting office mate and colleague for many years. Many other friends and colleagues contributed to a supportive and collegial environment in the office: Ivona Braviç, Beñat Mencia, Jan Behrends, Tycho Sikkenk, and Ryan-Rhys Griffiths all made my time spent in TCM and in the Maxwell Centre enjoyable. Importantly, I would like to thank Victor Jouffrey, Michael Rutter, and Gareth Conduit for mutually annoying each other, which has been entertaining and amusing to observe.

Finally, I would like to thank my parents for their constant and unconditional support throughout all of my PhD, during which they spent many hours listening, advising, and encouraging on so many occasions. More importantly, I would like to thank them for supporting me when I was younger, which provided me with the prerequisites needed to get accepted as a PhD candidate at the University of Cambridge.

Table of Contents

Nomenclature

Acronyms / Abbreviations

2D	Two dimensions or two-dimensional
3D	Three dimensions or three-dimensional
AC	Alternating current
AUC	Area under curve
DC	Direct current
DFT	Density-functional theory
DNEB	Doubly nudged elastic band
DNN	Deep neural network
FFLO	Fulde-Ferrell-Larkin-Ovchinnikov, used to refer to the work by these authors in Refs. [52, 76]
HLN	Hikami-Larkin-Nagaoka, used to refer to the work by these authors in Ref. [61]
LFL	Loss-function landscape
PEL	Potential-energy landscape
LJAT	Lennard-Jones Axilrod-Teller
LJAT_3	Three-atomic LJAT potential and geometry optimisation problem
LJAT19	Dataset generated for this work based on LJAT_3 geometry optimisation, see App. A.2
OPTDIG	Dataset of optical data of handwritten digits [2], taken from UCI Machine Learning Repository [46]

SARPES Spin and angular-resolved photoemission spectroscopy

SGD Stochastic gradient descent

TS Transition state

WINE Dataset of red and white wines with features being quality and 11 physiochemical tests [32], taken from UCI Machine Learning Repository [46]

Chapter 1

Introduction

Since the earliest attempts by physicists of the 19th century to explain the energy exchange and heat transfer between bodies, the notion of equilibrium has been a fundamental concept of thermodynamics – possibly *the most* fundamental concept [54]. It was perhaps unknown at the time of its inception in how many ways this concept could be interpreted beyond the study of the heat transfer of matter.

Today, the concept of equilibrium – or *thermodynamic* equilibrium, to be more precise – allows us to understand phenomena that are observed across a broad range of scientific problems. For example, chemists rely on the equality between forward and backward reaction rates in chemical equilibrium to determine the ratio of concentrations between reactants and products [72]. Quite similarly, the physics of semiconductors makes use of the requirement of constant electrochemical potential of a system in equilibrium to understand the behaviour of doped semiconductor devices [5]. But not only systems in equilibrium are of interest, in fact systems out of equilibrium, their relaxation, and even systems that cannot reach equilibrium, such as glasses and non-ergodic systems, provide physicists with a range of intriguing phenomena to study. This work investigates the equilibrium, non-equilibrium, and relaxation for a further two systems, which are distinctly different in nature.

The first study presented in this thesis investigates deep neural networks and their optimisation behaviour. In this context, the concept of thermodynamic equilibrium is taken further to assess the equilibration properties of systems that cannot directly be linked to any real physical system, but whose dynamics can nonetheless be described and studied by the adoption of methods from statistical mechanics.

Deep neural networks are perhaps the most relevant example from the toolbox of machine learning methods [42] and have over the past decade evolved to become a standard approach across many different fields of scientific research [93, 97, 132] and industry [101, 141]. Yet, how this method works or from where its outstanding performance originates has thus far been a mystery to a large extent [14, 123]. Technically, the training of deep neural networks is very similar to the fitting of a curve to a set of data points – something frequently encountered in physics – albeit with an extremely large array of tunable parameters as well as a high number of input and output variables. The task of adjusting this complex high-dimensional non-linear fitting function to a given set of data, a procedure referred to as either learning or optimisation, amounts to nothing more than a minimisation of a 'loss' or 'cost' function that estimates the discrepancy between curve and data.

As is known from both global optimisation of high-dimensional continuous functions as well as curve fitting with many tunable parameters, this task is far from trivial. Usually, basic optimisation methods that rely on gradient-based update procedures without the inclusion of a more sophisticated global navigation strategy easily end up trapped in local extrema. Yet, it comes as a surprise that stochastic-gradient descent – a method that serves as a foundation for many commonly employed optimisers – does exactly this and yet reaches deep neural network solutions that generalise well. From the perspective of physics, the problem of deep learning optimisation can be regarded as the motion of a physical system in a high-dimensional complex loss function, much like what is studied in the context of theoretical chemistry [145], where the thermodynamics and kinetics of molecules or solids are investigated through the study of the potential energy landscape that arises from inter-atomic interactions [4, 21, 25, 71, 83].

We shall use methods developed for the study of potential energy landscapes in molecular systems from the field of theoretical chemistry [145] and apply them to the loss function landscapes of deep neural networks, which yields useful conclusions for the dynamics that any optimiser based on the principles of stochastic-gradient descent would exhibit and hence also for the efficacy of deep learning optimisation.

The second study of this thesis analyses the decay of a non-equilibrium state in a solid-state system subject to a strong Rashba effect [23], which arises from spin-orbit coupling and a broken inversion symmetry. Such systems are more conventionally studied in the context of equilibrium and non-equilibrium statistical mechanics, yet phenomena novel to the field of solid-state physics are observed and reported.

Solid-state systems with strong Rashba coupling constitute a class of materials that have received much interest in recent years in the context of spintronics [153, 163] due to their ability to manipulate spin currents through the application of external electric fields in those systems [104] as well as in the context of non-centrosymmetric superconductivity, which can exhibit Majorana fermions [120, 121, 134] and is therefore of interest for fault-tolerant quantum computing [118]. Consequently, understanding non-equilibrium dynamics in those systems is of great relevance.

In this work, we investigate their low-temperature non-equilibrium properties and, in particular, the decay of chirality imbalances of charge carriers. In doing so, we aim to explain previously unreported non-equilibrium phenomena in low-temperature strongly Rashba-coupled systems, for which experimental findings are reported and discussed. This novel effect is most prominently realised in the normal and superconducting transport of GeTe, which, as we are able to demonstrate, cannot be understood or explained based on existing theoretical models. Those novel effects are not only relevant because their qualitative behaviour is unconventional, but also because the timescale over which the non-equilibrium effect relaxes turns out to be on the order of several minutes. This is far beyond the timescales that electronic carriers typically decay on, yet we present evidence that allows us to attribute the effect to electronic degrees of freedom of the system, which makes this observation intriguing. The generation and control of electronic non-equilibrium states is of great importance for both understanding condensed-matter systems as well as for the design of new devices relevant for technological applications. The feasibility of generating such non-equilibrium states hinges on a long relaxation timescale, which has sparked much interest in the search for long-lived electronic states in solid-state systems over the last few decades across different fields. For example, the endeavour to replace semiconductor-based technologies by spintronics devices has led to major interest in spin-relaxation timescales in spin-polarised systems [153]. Electronic momentum relaxes typically on timescales on the order of a picosecond, while spin-relaxation timescales can range up to microseconds at the longest [156, 163]. Consequently, the ability to attribute the experimentally observed long relaxation timescales in Rashba systems to an electronic non-equilibrium state would be an exciting result.

Following an assessment of relevant properties of the Rashba energy dispersion, we conclude that a non-equilibrium charge carrier distribution in the two Rashba bands with opposing chirality could be seen as a source for the slowly decaying non-equilibrium transport properties in the investigated samples. Consequently, we develop a theoretical framework based on chirality imbalances in Rashba systems, which capture the salient features of these experimental discoveries. This framework requires a heuristic approach to the modelling of carrier imbalances based on the well-known method of the relaxation-time approximation. To further test this framework with regards to the relaxation time, which is an important quantity to assess the applicability of the developed theory for the observed experimental effects, we present in-depth calculations that estimate the values of the relaxation time for different types of relaxation mechanisms. This study of the relaxation mechanisms allows us to gain a deeper understanding of chirality relaxation in low-temperature Rashba systems and verify whether the new framework based on Fermi-level imbalances can serve as a potential candidate to explain the novel non-equilibrium transport properties for GeTe observed in experiments.

This thesis is therefore split into Part I and Part II, where each part deals with one of the two main studies outlined above. Chap. 2 introduces the concept of deep neural network optimisation and explains the deep learning conundrum that we attempt to resolve. This is followed by an in-depth analysis of the loss-function landscape structure of deep neural networks using methods from molecular sciences in Chap. 3, which concludes Part I. The second part commences with Chap. 4, which introduces the required background details on Rashba systems and the employed non-equilibrium methods. Chap. 5 presents evidence of the above-mentioned novel non-equilibrium effect in low-temperature Rashba systems and develops a theoretical model based on slowly decaying chirality imbalances. Part II ends with Chap. 6, which again investigates non-equilibrium chirality charge carrier configuration in a two-dimensional free electron gas subject to strong Rashba coupling at low temperatures and estimates the relaxation timescales associated with different mechanisms that assist the decay of these non-equilibrium carrier configurations. Expressions for the relaxation-time constants are derived and quantitative estimates are compared to those observed in experiments. Finally, the work presented in this thesis is summarised in Chap. 7.

Part I

Equilibration in deep neural networks

Chapter 2

Introduction to neural networks and machine-learning optimisation

Over the last decade, a new class of methods has brought new inspiration and astonishing progress to many different fields of research: Machine Learning (ML) has enabled new scientific approaches that rely on the heuristic analysis of large amounts of data, resulting in new types of studies across a variety of subjects [14]. These methods make predictions, find patterns, and determine correlations from large datasets without the need for any knowledge of the hidden underlying principles that govern the relationships in those datasets. While these methods cannot be used to build new theoretical frameworks or deduce models from existing ones, they can however provide useful insight into where to look for undiscovered relationships or extract information from research data where human capacity is destined to fail due to the impracticably large amount of available data.

These new techniques have over the last few years produced surprising results and revolutionised research in some fields. As such, ML was recently used to learn the exchange-correlation energy functional in Kohn–Sham density functional theory from reference molecules [97]. Examples of applying ML to the identification of topological phases have also been reported [93]. Finally, a modelling of superconducting critical temperatures through the use of ML has also yielded intriguing results [132]. Clearly, the impact of ML in the field of condensed-matter physics will be just as big as in many other scientific fields, and the potential of ML is yet to be explored by future scientists. Therefore, even for condensed-matter theorists, it should be of interest to explore the capabilities of these tools to supplement their toolbox of research methods.

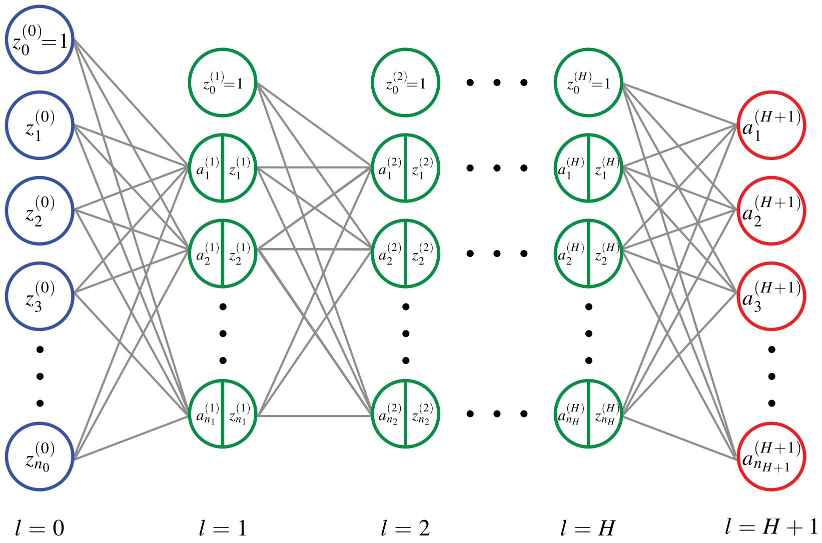
However, the work reported in this chapter and the following one does not attempt to use the capabilities of ML to make progress in condensed-matter theory. Rather, we take methods and tools developed from the physical sciences and apply them to ML methods to test and understand their functionality. Specifically, in this chapter we introduce the concept of a Deep Neural Network, which is a stereotypical prototype for an ML model. Such a model can be regarded as a large physical system with many degrees of freedom, its learning procedure as the dynamics of the system, and its loss function (i.e. the goodness of fit observed in the learning procedure) as its energy. We make this interpretation of deep neural networks more concrete in this chapter and apply methods of potential-energy landscape exploration from the field of theoretical chemistry in the next chapter.

2.1 Deep neural networks

In this section, we introduce the concept of a deep neural network (DNN) and explain how it is used in practice to learn patterns and correlations in large datasets in the context of supervised learning.

Fundamentally, a DNN is a fitting function with a large number of optimisable parameters. The function is constructed from an alternating concatenation of weighted linear combinations and non-linear transformations. This design is inspired by the way that neurons in the human brain operate, which gives this method its name. For a given entry of a dataset, the DNN takes a list of input features from that entry and computes a list of model output features from it. A linear combination of the input features is passed on to the next ‘layer’, where the non-linear transformation is performed, after which a linear combination is again passed on to the next layer. The function derived from this scheme is fitted to match the data from a specific dataset (also referred to as the ‘training’ dataset). The variable parameters of the fitting function are the weights of the linear combination, and their tuning can be accomplished using special optimisation algorithms. The latter is also referred to as the ‘learning’ or ‘training’ process of the DNN. The optimisation procedure relies on an estimate indicating the ‘goodness of fit’, often referred to as the loss function, such that the goal of the training process is to minimise this loss function. Once a DNN has been trained and a set of network parameters that sufficiently minimise the loss function has been found, the DNN can be used to predict the output features of unseen input features.

We note that the DNN discussed here can be regarded as a prime example from the vast variety of ML models. Other ML models exist as well, such as convolution neural networks, recurrent neural networks, and many other model types. These mainly differ in the model architecture employed, as well as the task to be performed on the data and hence the resulting optimisable loss function. Crucially, these models all use model architectures in the form of connected networks of linear transformations and non-linear activation functions. While the results derived in this work will differ for other model types, we yet expect many observations



Reprinted with permission from Ref. [142], P.C. Verpoort et al., Proc. National Academy of Sci. USA 117, 21857 (2020). © *2020 National Academy of Sciences.*

Fig. 2.1 Composition of a DNN. Input nodes are blue, output nodes are red, hidden nodes are green. A linear combination of the nodes from each layer is passed on to the next layer (except for the output layer, which is not passed on further). Each hidden node performs a non-linear transformation, as indicated by the vertical separation between the activations $a_i^{(l)}$ and the signals $z_j^{(l-1)}$ on these nodes. Bias nodes at the top of the input and hidden layers can be used to visualise the bias weights $\theta_i^{(l)}$, which can in this representation be incorporated into the weight matrices, $w_{ij}^{(l)}$. The entries of the weight matrices, $w_{ij}^{(l)}$, correspond to the links in grey connecting individual nodes but are not shown in this illustration for clarity.

reported here to be applicable more broadly for other model types as well. Finally, we note that there also exist other ML algorithms than the ones discussed here, including those that facilitate unsupervised learning or reinforcement learning, whose study is beyond the scope of this work.

We now start off by defining the architecture of the neural networks studied in this work in Sec. 2.1.1. Next in Sec. 2.1.2, we define the loss function and introduce the procedures commonly used for DNN optimisation.

2.1.1 Network architecture

The specifics of the DNN architecture considered in this work is shown in Fig. 2.1. We label layers using $l \in \{0,\dots,H+1\}$, with H being the number of hidden layers, $l=0$ the input layer, $l=H+1$ the output layer, and n_l the number of nodes in layer l. The activations $a_i^{(l)}$

are obtained from the signals $z_j^{(l-1)}$ using:

$$a_i^{(l)} = \sum_{j=1}^{n_{l-1}} w_{ij}^{(l)} z_j^{(l-1)} + \theta_i^{(l)} = \sum_{j=0}^{n_{l-1}} w_{ij}^{(l)} z_j^{(l-1)} \qquad \text{for } 1 \leq i \leq n_l\,, \tag{2.1}$$

where in the second step we have absorbed the bias weights into the link weight matrix by setting $w_{i0}^{(l)} = \theta_i^{(l)}$ and $z_0^{(l-1)} = 1$ for all l, giving these matrices an additional column and hence the shape $n_l \times (n_{l-1}+1)$. The number of variables is then given by $\nu = \sum_{l=1}^{H+1} n_l \times (n_{l-1}+1)$. The activations $a_i^{(l)}$ are converted into signals $z_i^{(l)}$ by applying the non-linear transformation function ϕ_l, such that

$$z_i^{(l)} = \phi_l(a_i^{(l)})\,. \tag{2.2}$$

While it would be interesting to determine how the choice of the activation function affects our results, we employ $\phi_l = \tanh$ for all layers l in the calculations performed in Chap. 3 and postpone further analysis to future work. We conjecture that similar results would be obtained from any another sufficiently smooth, monotonic activation function, yet this remains to be ascertained through the obtainment of further numerical evidence.

We note that the number of hidden layers of the network, denoted by the parameter H, is also referred to as the ‘deepness’ of the network. Networks with a large number of hidden layers are therefore often referred to as ‘deep’ networks, hence the name DNN. We note that the number of hidden layers in this work does not exceed[1] $H = 3$, yet we refer to these networks as ‘deep’ because we compare their performance with results for ‘shallow’ nets with $H = 1$. The number of nodes, n_l, in a layer l is referred to as the ‘wideness’ of the network. While the wideness can in principle change between layers (such as ‘bottleneck’ layers), we keep this number fixed for all hidden layers in the examples studied.

It is important to mention that the DNN architecture defined above features symmetry-related degeneracies. A degeneracy is given by two DNN models with different weights $w_{ij}^{(l)}$ that result in the same output features and the same loss-function value for any given set of input features. While degeneracies can occur accidentally[2], it is likely for these to be caused

[1] As explained further in Chap. 3, it would be desirable to go deeper than $H = 3$, yet this remains infeasible due to computational constraints for the moment. We nonetheless observe interesting shifts in the structure of the loss-function landscape when going from the shallow ($H = 1$) to the deep ($H \geq 2$) case in the models studied in this work and believe that these results have relevant implications to more sophisticated models despite the aforementioned limitations of the examples studied here.

[2] Note that we define a degeneracy as DNN models with both equal output features and loss-function value. Finding such a degeneracy is highly unlikely to occur for any DNN architecture and training dataset in practice. In contrast, degeneracies of just the loss-function value are much more likely to be present, especially for stationary points occurring at a high density, as further elaborated in App. A.3.

by symmetries. The most obvious symmetries in the DNN architecture used here are the permutational symmetries of the hidden nodes. It is clear that neither the final outputs of the DNN nor the resulting loss-function value should depend on how we label the nodes in each hidden layer. Consequently, a relabelling of the hidden nodes results in a simple permutation of the entries of the weight matrices, $w_{ij}^{(l)}$. The relabelling of nodes in hidden layer l with $1 \leq l \leq H$ is defined by the mapping

$$z_i^{(l)} \mapsto z_{\sigma(i)}^{(l)} \qquad \text{and} \qquad a_i^{(l)} \mapsto a_{\sigma(i)}^{(l)} \tag{2.3}$$

for all i with $1 \leq i \leq n_l$, where $\sigma \in S_{n_l}$ and S_n is the n-symmetric group, i.e. the group of all permutations of n elements. This relabelling necessitates the change of the weight matrices according to the following mappings:

$$w_{ij}^{(l)} \mapsto w_{\sigma(i)j}^{(l)} \qquad \forall\, 0 \leq i \leq n_l, 1 \leq j \leq n_{l-1} \tag{2.4}$$

and

$$w_{ij}^{(l+1)} \mapsto w_{i\sigma(j)}^{(l+1)} \qquad \forall\, 0 \leq i \leq n_{l+1}, 1 \leq j \leq n_l\,. \tag{2.5}$$

The first mapping, i.e. Eq. (2.4), becomes clear when inserting Eq. (2.3) into the definition of the $a_i^{(l)}$ in Eq. (2.1) and of the $z_i^{(l)}$ in Eq. (2.2). Consequently, a reordering of the hidden nodes requires a reordering of the corresponding rows of the weight matrix that connects the layer l with the preceding layer $l-1$. The second mapping, i.e. Eq. (2.5), is required in order to keep the values of the next layer unchanged. This is because

$$a_i^{(l+1)} = \sum_{j=0}^{n_l} w_{ij}^{(l+1)} z_{\sigma(j)}^{(l)} \tag{2.6}$$

will give a different result to Eq. (2.1), whereas

$$a_i^{(l+1)} = \sum_{j=0}^{n_l} w_{i\sigma(j)}^{(l+1)} z_{\sigma(j)}^{(l)} \tag{2.7}$$

gives the same result because the summation is commutative. Hence, the reordering of the hidden nodes in layer l additionally requires that the columns of the weight matrix that connects the layer l with the subsequent layer $l+1$ are reordered accordingly.

Note that the number of permutational degeneracies in a DNN equals to $\prod_{l=1}^{H} n_l!$, which grows quickly with the number of nodes in a hidden layer. Therefore, special care has to be taken in order to deal with these symmetries when applying the landscape-exploration tools in Chap. 3, as is further explained in Sec. 3.1.2. Finally, note that we refer to degenerate DNN models also as ‘isomers’ throughout the rest of this work due to the obvious correspondence between degenerate DNN models and molecular isomers in chemistry.

2.1.2 Network training

Having defined the DNN architecture in the previous section, we now turn to the issue of optimising this function in order to find a set of network parameters, $w_{ij}^{(l)}$, that optimise the match between the data and the network prediction.

For the training procedure of the network, we use a dataset with N_{data} entries, in which each entry is of the form $x_i^{(d)}$ for the input and $y_j^{(d)}$ for the output features, where the superscript index (d) numbers the data entries with $1 \leq d \leq N_{\text{data}}$ and the subscript indices i and j number, respectively, the input and output features with $1 \leq i \leq n_0$ and $1 \leq j \leq n_{H+1}$. In order to find a suitable optimisation procedure, we first need to define a measure that determines the ‘goodness’ of our fit. This is given by the so-called loss function L defined as

$$L(\mathbf{w}) = \sum_{d=1}^{N_{\text{data}}} L_d(\mathbf{w}) + L_{\text{reg}}(\mathbf{w}), \tag{2.8}$$

where $\mathbf{w} = (w^{(1)}, \ldots, w^{(H+1)})$ is the vector of all weight variables, $w_{ij}^{(l)}$, $L(\mathbf{w})$ is the full loss function, $L_d(\mathbf{w})$ is the loss for the data entry with index d, and $L_{\text{reg}}(\mathbf{w}) = \lambda\, ||\mathbf{w}||^2$ is an L^2 regularisation term[3] with regularisation parameter λ.

The loss $L_d(\mathbf{w})$ of each dataset entry d can take many forms. For example, a simple sum of squares can be defined as

$$L_d(\mathbf{w}) = \sum_{j=1}^{n_{H+1}} (y_j^{(d)} - a_i^{(H+1)})^2, \tag{2.9}$$

where the implicit dependence on $\mathbf{w}$ is encoded in the activations of the final layer, $a_i^{(H+1)}$, which can be computed by following the prescription outlined in the previous section. This loss function is typically used for regression problems, such as modelling the physical properties of a list of metallic compounds in a database of materials [141]. A loss function commonly used for classification problems is defined as

$$L_d(\mathbf{w}) = -\ln(p_{c(d)}) \qquad \text{with} \qquad p_i = e^{a_i^{(H+1)}} \Big/ \sum_{j=0}^{3} e^{a_j^{(H+1)}}, \tag{2.10}$$

[3] Note that $||\cdot||$ is the vector norm defined as $||\mathbf{x}|| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ for $\mathbf{x} \in \mathbb{R}^n$.

where $c(d)$ is the index of the known correct outcome for data item d obtained from the training data, and the p_i are the so-called ‘soft-max’ probabilities. This softmax loss function first converts the activations of the final output layer, $a_i^{(H+1)}$, to the probabilities p_i, which range between 0 and 1. The loss of an entry is then the negative logarithm of the probability that the network predicts the correct outcome for this entry, such that a minimisation of the loss function results in maximisation of the predicted probabilities of the correct outcomes.

The classification loss function in Eq. (2.10) will be used throughout the rest of this work, where the relevant classification problems are explained further in Chap. 3. We note that analysis similar to ours applied to other models with different loss functions would be interesting, yet we leave this to future work. Given the general independence of the optimisation performance on the precise loss function across many applications of deep learning, we expect to find similar results for different loss functions, however this remains speculation and would require further numerical evidence for confirmation.

The fundamental aim of any training algorithm is to minimise the loss function by updating the weights $\mathbf{w}$. In all cases except for some simple examples it is either impossible or at least infeasible to find local minima of the fitting function analytically, so the approach taken is entirely of numerical nature. Moreover, in most practical examples it is neither achievable nor desirable to find the global minimum of the fitting function. In fact, as is discussed later in Chap. 3, minima that are particularly low in their loss value for the training dataset are in some cases less likely to perform well on unseen data (also referred to as the testing dataset) than minima with a slightly higher loss value. This effect is also sometimes referred to as ‘overfitting’.

Optimisation of a DNN and its loss function can be achieved in many ways. Reviews can be found in Refs. [116] and [125]. Among the most important ones are Gradient Descent, Stochastic Gradient Descent [19], Weight-Decaying Gradient Descent and its derivatives, the Conjugate Gradient method [137], and the Levenberg-Marquardt method [53], which are all gradient-based optimisation procedures. This means that these methods compute the gradient of the loss function and subsequently take a step in the downhill direction of the gradient. Other procedures include Simulated Annealing [1, 111, 113] and Particle-Swarm Optimisation [69], which are more general optimisation schemes and can operate without knowledge of the gradient or even without the requirement of a differentiable loss function.

We note that in this work, we make explicit reference to stochastic gradient descent (SGD), which is the most widely used optimisation algorithm employed for training DNNs due to its simplicity in implementation, its robustness, and its effectiveness across many different applications. However we note that many of the research questions regarding how an optimiser navigates the loss-function landscape that we formulate in Sec. 2.2 and that we attempt to answer in Chap. 3 are not specific to the SGD algorithm and will be equally relevant to other related gradient-based DNN optimisers.

To complete this section, we briefly summarise how a common implementation of SGD would minimise a loss function like the one defined above in Eq. (2.8). A full gradient-descent optimiser would follow the simple approach

$$\mathbf{w}' = \mathbf{w} - \eta\, \nabla_{\mathbf{w}} L(\mathbf{w}) \tag{2.11}$$

to define a new set of weight variables $\mathbf{w}'$, where η is the step size, which is often also referred to as the 'learning rate' in the context of DNN optimisation and which is usually chosen in a heuristic fashion for the update procedure to yield good optimisation performance. It has however proven useful for several reasons explained later in this chapter to work with the gradient of the loss of just a few data entries $L_d(\mathbf{w})$ instead of the full loss function $L(\mathbf{w})$, which is often referred to as mini-batch SGD[4]. This yields an optimisation procedure that randomly divides all N_{data} entries into batches $D_1, \ldots, D_q$ of no more than b_{s} entries, where b_{s} is called the 'batch size' and we define[5] $q = \frac{N_{\text{data}}}{b_{\text{s}}}$, which is the number of batches. After this division of the dataset into q batches of b_{s} entries, the update procedure

$$\mathbf{w}' = \mathbf{w} - \frac{\eta}{b_{\text{s}}} \left[\nabla_{\mathbf{w}} \sum_{d \in D_k} L_d(\mathbf{w}) + 2\lambda \mathbf{w} \right] \tag{2.12}$$

is applied for $k \in \{1, \ldots, q\}$, before a new random division of the data entries into batches is performed. Note that the second term proportional to $2\lambda\mathbf{w}$ stems from the regularisation term introduced above in Eq. (2.8). The completion of such a cycle of division into batches and update of weights is also referred to as an 'epoch'. Typically, the batch size b_{s} is chosen such that the computation of $\nabla_{\mathbf{w}} L_d(\mathbf{w})$ in Eq. (2.12) can be parallelised (or vectorised, in the case of GPU-based computation) efficiently over d. While the original motivation to work with the mini-batch gradient instead of the full gradient was to reduce computational cost in the estimation of the loss gradient, it becomes apparent in the next section why this procedure in fact provides desirable characteristics for global loss-function optimisation. Mini-batch SGD adds a form of stochastic noise to the evaluation of the loss function and its gradient, which prevents trapping of the optimiser in local minima and therefore eases optimisation in a non-convex loss function.

[4] While many textbooks use the term SGD strictly to refer to mini-batch SGD with batch size $b_{\text{s}} = 1$, it has become customary in the ML community to use the term SGD to also refer to procedures with $b_{\text{s}} > 1$.

[5] Here we assume that N_{data} is an integer multiple of b_{s} for the sake of simplicity.

Finally, we note that the gradient of the loss function with respect to the weights, $\nabla_{\mathbf{w}} L_d(\mathbf{w})$, can be obtained while computing the loss function itself through a method referred to as 'back-propagation'. The computation of the second-order derivatives of $L_d(\mathbf{w})$ (also known as the Hessian matrix) is less common in the ML community, as it is usually not required for most gradient-based optimisers and requires more involved calculations. The knowledge of the second-order derivatives is however crucial to the work explained in the next chapter, so that a computation of both first and second-order derivatives of the DNN loss function, are outlined in App. A.1.

We note that knowledge of the second-order derivative is primarily useful for Newton-like minimisation, which we make heavy use of in the next chapter. Such minimisation techniques are however not usually employed in ML due to the significantly higher computational cost incurring in the calculation of the second-order derivatives and due to only minor performance gains for optimisation. It should be borne in mind that the target of optimisation is just minimisation of the loss function and not the identification of a stationary point. This is why the second-order derivative is relevant to the studies presented in this work but is often disregarded in the field of ML, where only the gradient is considered. We also note that the second-order derivative enables an analysis of the eigenfrequency spectrum for stationary points. While some analysis of the correlation between the Hessian eigenfrequencies and the model generalisability is presented in Sec. 3.2.3, we hope that future work will look more carefully at the number of soft and hard directions of minima, the resulting estimate of the basin volume, the precise basin volume obtained from thermodynamic integration, and how all that correlates with goodness of fit and generalisability, where analysis of all these aspects crucially depends on the calculation of the second-order derivatives.

2.2 The deep-learning conundrum

While the capabilities of DNNs to learn patterns and predict outcomes for unseen data has been highly successful and continues to be exploited in new emerging fields of industry and research, it is still unclear why common training procedures, especially those relying on SGD or other gradient-descent optimisation methods, can succeed in locating locally optimal points in the high-dimensional space of weight variables that sufficiently minimise the loss function. Without the evidence of the effectiveness of DNN optimisation using SGD based on a large number of highly successful practical examples, one would naively expect a gradient-descent optimiser to end up trapped in the catchment basins of local minima, and that transcending the barriers that separate these basins in order to sustain a global downhill trajectory is unlikely to be yielded by simple methods such as SGD [123].

We believe that this question can at least partly be answered by better understanding the structure of the underlying loss-function landscape (LFL) [12, 36, 107, 151], i.e. the high-dimensional hypersurface created by the loss value as a function of the weight parameters, $L(\mathbf{w})$, as defined in Eq. (2.8). One of the most important features of a LFL is its connectivity, determined by the barrier heights separating local minima, which we investigate in the next chapter. In the following, we make use of the analogy between the LFL in DNNs and the potential-energy landscape (PEL) in molecules and solids, and we therefore apply methods developed in the context of PEL exploration to the analysis of LFLs in DNNs. This allows us to gain a better understanding for the relevance of the structure of the LFL for DNN optimisation. In molecular sciences, the PEL is defined by the potential energy of a molecule or solid-state system as a function of the coordinates of each atom, where each local minimum corresponds to a molecular isomer or frozen solid-state structure, and the pathways between these minima correspond to atomic rearrangements.

Gradient-based optimisation of a DNN relies on two ingredients: (1) minimisation, i.e. moving in the direction of the downhill gradient to minimise the loss-function value, and (2) a mechanism that ensures that the minimiser does not converge to high-lying local minima, or ends up trapped in the catchment basins of such minima. This mechanism can manifest in different ways: In stochastic gradient descent (SGD), noise is provided by using only a fraction of all samples to compute the loss and its gradient[6], whereas simple gradient descent with stochastic noise has its noise added artificially on top. In basin-hopping [29, 80, 81, 148], a powerful global-optimisation technique employed in Chap. 3, random displacements in the parameter space are made to hop between the catchment basins of local minima. In physical systems, the forces driving the motion towards lower energy are given by the gradient of the energy, whereas noise to overcome barriers arises from thermal fluctuations[7].

It intuitively makes sense that a combination of (1) gradient-based local minimisation and (2) an 'untrapping' strategy can yield successful (sometimes even global) optimisation of a function, yet this is not guaranteed to work. For example, there exist many physical systems whose energy landscapes display a hierarchical structure, where a vast number of low-lying minima exist, separated by high barriers [39–41, 91]. The consequence is that those

[6] We note that the 'noise' introduced through mini-batch SGD is of course very different in its nature from pure stochastic noise [161] and, moreover, depends strongly on the dataset employed. For instance, note that the noise introduced by mini-batch SGD is small when all entries of a dataset are similar and is big when there is large variation between dataset entries. Moreover the strength of this noise also depends on the learning rate η, as for small learning rate an epoch will have passed before the optimiser has made any large steps in parameter space, whereas for huge learning rate the gradient would be entirely replaced by the gradient for just a subset of the data. Details on the relationship between the learning rate η, the batch size b_{s}, and the amount of training data N can be found in Refs. [130, 131].

[7] Notably, this is a fully classical picture that does not take into account the quantum nature of molecular dynamics, yet this simplified picture will be sufficient for our comparison with DNN optimisation.

systems exhibit glassy dynamics across a large temperature range. In such systems, thermal fluctuations are unable to untrap the system and guide it to a low-lying state. (Basin-hopping, on the contrary, has been shown to work even for glassy landscapes, and we employ it to explore the LFL in the next chapter.)

Given the complexity of a DNN and the number of locally optimal solutions (as determined in the next chapter), it is reasonable to expect high barriers separating those solutions and a complex loss-function landscape that is difficult to navigate. This structure would make minimisation through 'simple' optimisation procedures such as SGD infeasible, given that SGD untraps itself through noise in a similar fashion to thermal fluctuations of physical systems, as is evident from previous work by Zhang et al. [161] that maps the dynamics of SGD optimisation to Langevin-type equations of motion. Practical experience from SGD optimisation of DNNs however indicates the opposite: it is usually easy to find a set of parameters η and b_s that succeed in navigating the LFL and quickly locating low-lying points in the landscape[8].

To clarify this point further, we present numerical evidence of the trajectory of SGD in the LFL of a DNN. To achieve good optimisation with SGD, an appropriate choice of parameters (in particular batch size b_s and learning rate η) is key. In Fig. 2.2, we report the loss L during training as a function of epochs for varying values[9] of b_s (the training dataset employed is the one from the LJAT19-3HL-2000 example from the next chapter, but this is of minor relevance here). When the batch size is too small, and hence the noise too big, SGD becomes essentially a random walk in the landscape. Optimisation to low-lying minima is not possible, and instead the optimiser oscillates at loss values far above the global minimum. When the batch size is chosen too large, and hence there is insufficient noise, the optimiser gets trapped in local minima and does not reach low-lying solutions. The ideal parameterisation lies somewhere in between those two cases. These observations are widely known, yet understanding why this is possible given the high complexity of the LFL of a DNN remains a puzzle, which the study presented in the next chapter aims to uncover.

Another important question of DNN optimisation is why SGD finds solutions that generalise well. One might ask how well a low training loss value might correlate with good performance on unseen testing data. This is particularly relevant when overfitting

[8] Note that the aim of optimisation in ML is not to identify stationary points. While gradient-based minimisation pursues the target of minimising the gradient, which would result in the discovery of stationary points, the ultimate goal is simply to find points in the LFL whose loss value is low, hence indicating a good fit of the model with the training data.

[9] Note that η is kept fixed in the presented examples. We could have alternatively varied η and kept b_s fixed. In fact Refs. [130, 131] suggest that it might even be more advisable to look at the "noise scale", defined as $g = \eta\left(\frac{N}{b_\mathrm{s}} - 1\right)$, when comparing the different cases shown in Fig. 2.2, which is however beyond the scope of this work.

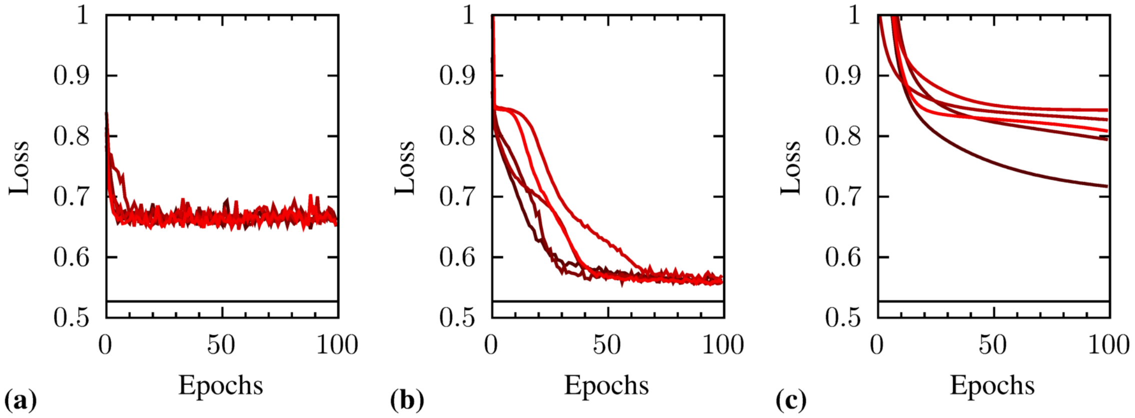


Fig. 2.2 Loss function L plotted as a function of training epochs for SGD optimisation of a DNN model. Each line in the plot corresponds to a different random initialisation. The black line at the bottom of each plot is the loss of the lowest minimum obtained from basin-hopping. This procedure explores the LFL of a DNN architecture with $N_{\text{data}} = 2000$ entries and $H = 3$ hidden layers using the LJAT19 dataset introduced in Chap. 3 and learning rate $\eta = 0.1$. **(a)** $b_{\text{s}} = 1$. The batch size is too small (for the given step size), and, consequently, the optimisation is poor. **(b)** $b_{\text{s}} = 100$. The noise is not too big so that efficient local downhill optimisation is possible, equally the noise is not too small, allowing the optimiser to overcome barriers between minima. **(c)** $b_{\text{s}} = 2000$. The batch size is too big, hence the optimiser is trapped in local minima and does not reach low-lying solutions.

occurs, which we can expect to be the case in most practical examples, as it has proven to be successful to choose DNNs with numbers of variables, ν, of the same order of magnitude as the amount of training data, N_{data}. It has been shown that SGD optimisers prefer wider minima with a low curvature over ones with a high curvature [161], and improved performance of DNNs has been achieved by further biasing the optimisation towards such solutions [10, 26]. We therefore analyse the correlation between training and testing loss values in the next chapter, after having successfully obtained an accurate and comprehensive database of all local minima of the LFL.

The present work is not the first to study the LFLs of DNNs. Previous works either developed analytical models that investigate the LFL for simple DNNs, usually by assuming random data and making heavy approximations, or numerically studied the LFL of concrete DNN training examples, in which however complete and accurately converged databases of local minima and transition states are usually not obtained.

On the analytical side, Ref. [89] described the dynamics of optimising a DNN with $H = 1$ using SGD and proved the 'convergence of SGD to a near-global optimum' by studying the partial differential equation governing those dynamics for this simple example.

In Ref. [31], DNNs with linear rectifiers as activation functions, with random training data, and without bias nodes are mapped to spin-glass models with spherical constraints. The authors use random-matrix theory to gain insight into the distribution of minima of the LFL. Unfortunately, their study cannot provide any details about barriers separating those minima. Nonetheless, it reveals that local optima are 'located in a well-defined band lower-bounded by the global minimum'. This intriguing result can at least partly be confirmed through our results reported in the next chapter. However, it should be noted that their mapping of a DNN to a spin-glass model involved a number of significant approximations.

On the numerical side, geometric properties of LFL minima have been reported, most notably their width and how this relates to the generalisability of a model [62, 63, 65, 70]. The vast majority of numerical LFL studies however do not explore the connectivity of minima. Some exceptions exist though, including studies of the LFL for single hidden-layer networks [11, 12, 30, 36, 37], which use a formalism similar to the one we employ in the next chapter. Essentially, the work presented here can be regarded as an extension of this existing work on the LFL of single-hidden layer networks to multi-layer networks. Moreover, Ref. [79] developed a method for visualising, quantifying and comparing LFL complexity that reveal astonishing results for skip-connection networks, i.e. networks that feature additional connections between non-neighbouring layers to promote better learning capabilities and avoid the vanishing-gradient problem in optimisation. Finally, a very intriguing study by Draxler et al. in Ref. [45] observed that pairs of minima in the LFL of DNNs are connected by low-barrier minimal-loss pathways, which encounter much lower barriers than the direct interpolation path between those two minima.

Finally, a combined analytic study of the minima geometry in LFLs of DNNs attributes the effectiveness of SGD to the different sizes of the basins of attraction of 'good' and 'bad' minima [155]. While these results are interesting and are revisited in the next chapter, where we also analyse the correlation between Hessian matrix eigenvalues and generalisability of minima, that study does not provide insights into the barrier heights or the navigation of the LFL by a SGD optimiser.

While all these studies, both analytical and numerical, have contributed significantly to our understanding of the structure of LFLs of DNNs, they can only report isolated results on some aspects and lack a global view that connects all these individual observations. We aim to deliver this missing global perspective of the LFLs of DNNs in the next chapter.

Having completed an introduction to DNNs, their optimisation through SGD, and the importance of the structure of their LFL in this chapter, we proceed by analysing the LFL for a set of DNN examples in the next chapter. We employ techniques from global optimisation for energy landscapes in molecular sciences to resolve the structure of the LFL, which allows us to gain insights into the number of training minima, their distribution of loss values, the height of the barriers connecting them, and their Hessian eigenvalues. We use these results to gain deeper understanding of the structure of the LFL and, consequently, why SGD succeeds in navigating it.

Chapter 3

Loss-function landscapes for deep neural networks

In this chapter, we study the structure of the LFL of DNNs, and we do so by determining the number of local minima and transition states in these LFLs, the barriers separating minima, and the correlation between minima geometry and model generalisability. We employ the computational energy-landscape exploration techniques and the disconnectivity-graph formalism originally developed for the study of PELs in molecular science [15, 145, 150]. We perform this analysis for varying numbers of hidden layers, H, and the amount of training data, N_{data}.

The networks studied here are sufficiently small that we can can efficiently converge the databases of minima and transition states. We note that these networks comprise relatively small architectures compared to the ones commonly used in typical ML applications. However, we conjecture that the fundamental observations in these small networks, which feature up to 5×10^5 local minima and 10^6 transition states, can be generalised to larger and more complex networks, where the dimensionality of the LFL it too big to enumerate all minima or even converge a single local minimum accurately.

The results obtained from the study presented in the following sections, which we summarise further at the end of this chapter, reveal that the LFLs of DNNs in the shallow ($H = 1$) or data-abundant ($N_{\text{data}} \gg 1000$) limits feature a single-funnelled structure, in which the downhill barriers are small. In contrast, for a multi-layer network ($H \geq 2$) and for little training data ($N_{\text{data}} \lesssim 1000$) the LFL is characterised by a large number of minima of similar loss values that are separated by low barriers. Both of these landscapes are different from the hierarchical landscapes observed in structural glasses, which helps us understand why procedures commonly employed by the ML community can navigate the LFL successfully and reach low-lying solutions.

This chapter is structured as follows. The employed methods are outlined in Sec. 3.1, where we introduce the datasets used for training and the landscape-exploration tools borrowed from the molecular sciences. This is followed by a presentation of the results in Sec. 3.2, where we visualise the resolved LFL structures using disconnectivity graphs and present statistics on the number of minima, number of transition states, their loss values, barrier heights, and the correlation to minima geometry. Finally, we conclude with a summary and an interpretation of our results in Sec. 3.3.

3.1 Methods

3.1.1 Training and testing data

We analyse the LFLs for three distinct datasets: LJAT19 [142], OPTDIG [2], and WINE [32], as summarised below.

The LJAT19 dataset was generated specifically for the assessment of DNN loss landscapes. It is based on the outcome predictions of geometry optimisation of the LJAT$_3$ problem, which was used in previous work [11, 12, 30, 37]; details of its creation are reported in App. A.2. Readers primarily interested in LFLs need simply note that this is a four-fold classification problem with the correct outcome being fully determined by the three inputs. Here, we only employ two of these inputs for training and testing to make this problem harder. This benchmark is appealing because we can generate arbitrary amounts of training and testing data, and it has practical importance in chemistry, where calculations of molecular configuration volumes and densities of states are of interest. In the case of LJAT19, we assess the performance of both a training and a testing dataset. These datasets are obtained by generating 200 000 entries from geometry optimisation and splitting the data up into two random subsets of $N_{\text{data}} = 100\,000$ data points, which are then referred to as the training and testing data, respectively. When fewer data are employed in training, i.e. $N_{\text{data}} < 100\,000$, we use the first N_{data} entries of the complete training dataset.

OPTDIG is a set of optical data for handwritten digits with the target being digit recognition/classification. This resource is similar to the widely known MNIST dataset, but with inputs of size 8x8. The WINE dataset is a list of red and white wines, with the target being the classification of their quality as ‘good’, ‘medium’, or ‘bad’ based on 11 physicochemical tests (such as acidity, sulphates, alcohol, etc). Both OPTDIG and WINE are ‘real-life’ data, and were obtained from the UCI Machine Learning Repository [46].

We primarily focus our studies on the LJAT19 dataset because we can generate as much data as is needed for benchmarking, which allows us to produce disconnectivity graphs with up to $N_{\text{data}} = 100\,000$ training data. In contrast, the WINE and OPTDIG datasets only have up to $N_{\text{data}} = 5000$ entries, so we use these to validate and confirm our principal conclusions. The computational cost of our methods grows quickly with the dimensionality of the LFL, and hence the datasets we study are a compromise between complexity and feasibility.

The performance of any point in the LFL, defined by a set of weight variables $\mathbf{w}$, can be quantified using standard Area Under Curve (AUC) receiver operating characteristic metrics [57]. AUC values range between 0 and 1, where random classification yields an AUC value of 0.5, and a perfect prediction yields 1. In addition, we visualise the prediction outcomes of points in the landscape of the LJAT19 dataset by colouring a representative subspace of the inputs according to the classification indices (see caption of Fig. 3.2, and App. A.2 for further details).

The L^2 regularisation parameter used for training and testing throughout the rest of this work is $\lambda = 10^{-4}$. Decreasing this value would make the landscape more complex and increase the number of stationary points, while an increase would result in fewer local minima and a simpler landscape structure. While a study of the dependence of the LFL structure as a function of λ would be interesting, we restrict ourselves to a constant value here and postpone the investigation of the dependence of the LFL on the value of λ to future work.

3.1.2 Exploring the loss-function landscape

As mentioned before, this chapter makes heavy use of geometry optimisation tools developed for the analysis of PELs in molecular sciences and applies them to the exploration and visualisation of LFLs of DNNs. Several extensive reviews of this methodology exist [11, 67, 145–147], hence the introduction to these methods presented below is intentionally kept concise.

The exploration of the landscape can be separated into two main steps. First a (likely incomplete) list of minima is obtained through global optimisation using basin-hopping (BH). In a second step, approximate steepest-descent paths connecting two local minima pairwise are determined, from which the transition-states (TS) and hence the connectivity of the landscape can be obtained.

BH is a simple yet powerful technique for global optimisation of high-dimensional non-convex functions [29, 80, 81, 148]. It starts off with performing a gradient-based minimisation to identify a zero-gradient stationary point. The discovered stationary point is typically the lowest point of the basin of attraction of the starting position, although this cannot be guaranteed and depends on the step-size. After this minimisation has succeeded and a stationary point has been identified, a random displacement in variable space is taken, and the function value for the new candidate position is evaluated. The new candidate position is adopted according to the satisfaction of the following Metropolis-type condition: the displacement step is always accepted if the new resulting function value is lower than the current one and is otherwise accepted with probability $P(L,L') = \exp(\beta\,(L'-L))$, where L is the function value at the current position, L' is the function value at the new candidate position, and β is the inverse fictitious temperature.

The parameter β should ideally be chosen to be roughly on the order of magnitude of the average barrier height that has to be overcome. In physical systems, choosing this parameter is guided by knowledge of physical quantities, which can in some cases be determined from experiments. In this work, β was chosen in a heuristic way that allowed an efficient discovery of the low-lying minima of the landscape. Choosing a β that is too large will discover lots of new minima but will struggle to identify low-loss solutions, whereas choosing a β that is too low will end up trapped in the catchment basins of just a small sub-funnel of wider landscape (similar to what is shown and discussed in Fig. 2.2). A good choice of β will discover the global or at least a near-global minimum, revisit this minimum repeatedly, and not identify any new lower minima after reaching convergence following a large number of steps. We note that this approach does not guarantee finding the global minimum. However, we are more interested in the relative organisation of minima and macroscopic structure, for which knowledge of the global minimum is not needed. Moreover, the minima discovered by BH only act as a starting point for a more extensive analysis using TS searches.

In summary, BH combines approaches from gradient-based methods for continuous function optimisation with a global combinatorial optimisation strategy. It has been shown that BH is able to navigate complex landscapes successfully, escape from basins of locally optimal minima in order to maintain a global downhill trajectory, and often locate the global minimum even in multi-funnelled landscapes [27] or the PELs of structural glasses [39–41, 102, 103]. Importantly, BH is not subject to the exponential slow-down suggested by spin-glass models [31].

The local minima obtained from BH are not sufficient to understand the structure of the LFL. In order to establish a complete picture of the landscape, the connectivity of the minima has to be determined as well. This step is accomplished by identifying the TSs connecting the local minima obtained from BH. TSs are defined as stationary points (i.e. points with vanishing gradient), for which the number of negative eigenvalues of the Hessian matrix (also referred to as the Hessian index) is precisely one. Hence, following the two steepest-decent paths from a TS lead to the two minima it is connecting, while the curvature remains positive in all other directions. The Murrell-Laidler theorem [96] guarantees that TSs rigorously define the lowest-barrier paths connecting the catchment basins of local minima. Higher-index saddles (i.e. stationary points with Hessian index greater than one) exist as well, but they are not relevant to the study of the lowest barriers separating local minima. We note that, arguably, higher-order stationary points might play an important role in DNN optimisation because commonly employed optimisers may erroneously converge to such points. However this problem is beyond the scope of this work. For that reason, the term ‘barriers separating local minima’ used in this work refers to the highest point of the minimal-loss pathways between minima and is the lowest possible loss value any optimisation procedure, including ML optimisers such as SGD, must surmount to pass between the catchment basins of two connected local minima [90, 145].

Local TS searches can be performed using eigenvector-following procedures [95, 159], which are an extension of the Newton-Raphson method [145]. This procedure employs a modified Newton-Raphson step that moves uphill in one and downhill in all other Hessian eigenvector directions and converges to the standard Newton-Raphson step in the vicinity of the stationary point. While this method is useful for converging potential candidates for TSs, it is ineffective for global TS searches.

To identify TS candidates, a doubly nudged [138, 139] elastic band [59, 60] (DNEB) approach is used here. This method places a fixed number of ‘beads’ between a starting and end point in the landscape and optimises their position using a fictitious force consisting of the loss gradient in all directions orthogonal to the tangent vector plus a spring force between the beads. The optimisation of this elastic band of beads can be performed by integrating the Newtonian equation of motion or, as is done in this work, by using more advanced optimisation methods, such as a limited-memory quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (LBFGS) routine [22, 51, 56, 124]. This procedure yields a series of approximate intermediate positions on the steepest-descent paths, of which any one can be used as an input for hybrid eigenvector-following in order to obtain a TS.

When this approach is taken for two given minima in the landscape, there is no guarantee that there exists a single steepest-descent path that connects the initial to the final point. Rather, following a DNEB procedure that yields a TS candidate and a convergence of the candidate to a true TS of the LFL, the correct end points that the steepest-descent path leads to downhill from the identified TS have to be determined. This procedure may or may not result in the initial minima used in the DNEB approach and often discovers new minima, thus extending the database of discovered minima. Strategies for selecting a pair of minima for connection using the DNEB approach in order to efficiently build the connectivity network have been developed, such as the missing connection algorithm [24], which is also employed in the present work.

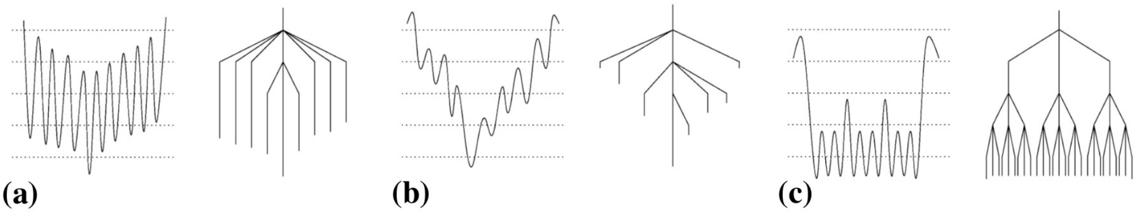
The combined application of these three methods – BH for discovering minima, DNEB searches between minima to identify TS candidates, and hybrid eigenvector-following for converging those TS candidates – produces a database of connected minima and TSs. This database can be regarded as the analogue of a kinetic transition network in molecular sciences, where the techniques outlined above have been applied extensively in previous works. Standard procedures developed for expanding those databases [11, 67, 145–147] can equally be applied to the study of LFLs. As already reported in previous work in Ref. [88], the fact that the Hessian eigenvalues can be spread across several orders of magnitude between different directions makes the convergence of pathways harder for LFL compared to PELs. Moreover, as established in Ref. [88], a suitable method for dealing with symmetry-induced degeneracies of local minima lies in a tight convergence of their loss values and an identification of equal representations of the same minimum purely based on their loss value. This approach is required because the degeneracies induced by the permutational symmetry of the hidden nodes, which are unlike the ones commonly found in molecules, are hard to treat with a distance metric, at least when the number of hidden layers exceeds one[1]. We note however that this simplified approach to permutational degeneracies results in deficiencies in our results because different minima with similar loss values may falsely be identified as equal representations of the same minimum.

[1] While in this work, we detect different representations of identical minima with permuted weights through the similarity of their loss values, a more robust approach would be to test for their proximity in weight space. This approach, however, requires determining the minimal distance between two given points in the LFL under consideration of all possible symmetries.

In the molecular sciences, this calculation involves finding the minimal distance between two isomers for all combinations of permutations of atoms of the same species along with translational and rotational rearrangements. This calculation is achievable because translational and rotational symmetries are continuous symmetries, and because the determination of the specific permutation of equal atomic species that minimises the distances between two isomers can be mapped to a linear assignment problem. This setup does not naturally carry over to DNNs, where the symmetries are very different in nature.

The symmetries in DNNs are mainly given by permutation of hidden nodes inside any hidden layer. Trying out all permutations in order to find the one that minimises the distance in weight space is NP-hard and unfeasible even for moderately large networks; note that the number of permutations scales as $n!$ with n being the number of hidden nodes to permute, which results in approximately 3×10^6 permutations for networks with 10 hidden nodes. For one hidden layer, finding the optimal permutational alignment is a linear assignment problem and can be solved in polynomial time using the Hungarian method [75]. For $H = 2$, this is already NP-hard, although approximating algorithms have been reported [33]. For $H > 2$, no efficient algorithms are currently known.

In practice it may suffice to employ a much simpler scheme that does not rely on computing the minimal distance but permuting all hidden nodes in each layer so that they are brought into a normal ordering with a strictly increasing bias weight. The success of this method relies on the absence of a degeneracy in the bias weights and does not have the desirable properties of a distance metric unless the two considered points in the LFL are sufficiently close. Yet, this approach could help with the identification of different representations of points under all permutations in the LFL, which is however beyond the scope of this work.



Reprinted with permission from Ref. [150], Springer Nature: Archetypal energy landscapes by David J. Wales et al. (1998). © *1998, Macmillan Magazines Ltd.*

Fig. 3.1 Examples of one-dimensional non-convex functions and their corresponding disconnectivity graphs, reprinted from Ref. [150]. Original caption: "Pictorial correspondence between the potential-energy surface and the disconnectivity graph for three different energy landscapes, following Becker and Karplus [15]. **(a)** The 'weeping willow' results from a gentle funnel with large barriers. **(b)** The 'palm tree' results from a steeper funnel with lower barriers. **(c)** The 'banyan tree' results from a rough landscape."

We note that a nudged elastic-band approach was previously employed in the study of barrier heights in neural network LFLs in Ref. [45]. This work, however, only established approximate TS candidates using the nudged elastic band method and did not converge these candidates, as is done in the present work, using hybrid eigenvector-following. A DNEB interpolation only yields a series of discrete images that are neither TSs, nor does their described pathway necessarily correspond to a single steepest-descent path. Consequently, additional stationary points may be skipped, resulting in missed minima, TSs, and, in the worst case, an incorrect estimation of the barrier height. It is therefore necessary to first refine the TS candidates, followed by a calculation of an approximate steepest-descent pathway to identify the connected minima.

Finally, we report our results of the LFL of DNNs using disconnectivity graphs, a form of visualisation previously used to present results for PELs [15, 150]. For readers unfamiliar with disconnectivity graphs, Fig. 3.1 presents reprints from Ref. [150] that illustrate the correspondence between LFLs and disconnectivity graphs for simple one-dimensional non-convex functions. Disconnectivity graphs are used to visualise the loss values of minima, along with the height of the barriers separating them. Each graph consists of a tree diagram, where every branch end point corresponds to a local minimum. The loss value is on the vertical axis, and so the height of the end point of a branch represents the loss value of the corresponding minimum. A superbasin analysis of all minima is performed at thresholds of regular loss value intervals, where all minima that can be connected via TSs below the threshold are grouped together. The branches of a group of connected minima are merged together into a single branch at each threshold, and the branches are positioned on the horizontal axis in a way that avoids branch crossings.

3.2 Results

First, we study the LFL of DNNs using the LJAT19 dataset. We characterise the LFLs as a function of training data ($N_{\text{data}} \in \{100, 1000, 2000, 10\,000, 100\,000\}$) and number of hidden layers ($H \in \{1,2,3\}$). We choose $n_l \in \{10,5,4\}$ for each hidden layer l, which yields an approximately equal number of weight variables ($\nu \in \{74,69,72\}$). The number of input and output features are $n_0 = 2$ and $n_{H+1} = 4$, respectively. For each of these cases, the procedure outlined in Sec. 3.1.2 is performed to calculate databases of local minima along with TSs connecting them pairwise. We report the number of minima and TSs, visualise the LFL using disconnectivity graphs, and analyse the correlation between train error, test error, and minima geometry. We reiterate that the above values of ν are small compared to those found in typical applications, yet they exhibit LFLs that feature sufficiently many local minima to identify trends, and for which optimisation procedures can be converged accurately.

We note that the lowest minimum discovered from BH remains the lowest in all cases on expanding the database, except for $H = 3$ and $N_{\text{data}} = 100$. This result indicates that the landscapes are sufficiently easy to optimise with BH such that we can uncover the fundamental organisation of the landscape independently of the exact details of the optimisation protocol.

After our study of the LJAT19 dataset we confirm that our key results also apply to the OPTDIG ($N_{\text{data}} \in \{1500, 5000\}$, $H \in \{1,3\}$) and WINE ($N_{\text{data}} = 1500$, $H \in \{1,3\}$) datasets, where we simply report the disconnectivity graphs for comparison.

We stress that the employed procedure, which combines BH global optimisation with TS searches using DNEB approaches and hybrid eigenvector following, is in principle capable of discovering all minima and TSs in the LFL, provided the process is continued for long enough for the database to be converged sufficiently. However, there exists a deficiency in our procedure due to the potential miss-identification of minima as permutational isomers. This problem is further explained and its impact investigated in App. A.3. In summary, in some cases of the LJAT19 dataset, in which the minima density becomes high enough such that the loss difference between minima becomes less than the loss difference tolerance used in the identification of permutational isomers, minima are incorrectly treated as permuted representations of known minima. While we interpret this mishandling of some of the minima in LFLs in those cases as of only minor importance for the overarching trends reported in the following sections (except perhaps for the total number of minima and TSs reported in Sec. 3.2.1), we stress that future work ought to revisit the results presented here using more advanced procedures for treating the permutational degeneracy problem, including methods that accurately estimate the geometric distance in DNN weight space. We also note that this aspect is not of relevance to the analysis done using the OPTDIG or WINE datasets, where only a small fraction of all minima and TSs in the respective LFLs were discovered in order

Table 3.1 Number of local minima and TSs for DNNs with varying number of hidden layers, H, and training data, N_{data}, using the LJAT19 dataset.

	$H=1$		$H=2$		$H=3$	
N_{data}	n_{min}	n_{ts}	n_{min}	n_{ts}	n_{min}	n_{ts}
100	649	10 426	299 484	508 471	585 730	1 126 877
1000	7	33	13 336	41 837	85 150	263 615
2000	5	21	487	1583	3027	35 363
10000	7	43	49	393	150	804
100000	5	11	33	200	23	96

to plot a disconnectivity graph, hence showing just a selection of the features determining the LFL structure.

All the results are labelled as 'DATASET-#HL-#', where the two digits indicate, respectively, the number of hidden layers, H, and the amount of training data, N_{data}.

3.2.1 Number of minima

First, the clearest and most unsurprising trend in the LFLs is easily summarised: the number of local minima, n_{min}, and the number of transition states, n_{ts}, for which data is reported in Tab. 3.1, decrease rapidly with increasing N_{data} (until N_{data} reaches a critically low value, after which oscillations of n_{min} and n_{ts} start to occur). All the reported values are lower bounds, as a full convergence of the procedure outlined in Sec. 3.1.2 cannot be guaranteed, partly due to the miss-identification problem for symmetry-unrelated minima with similar loss values, as discussed at the beginning of Sec. 3.2 and in App. A.3. The largest databases for $N_{\text{data}} = 100$ and $H \in \{2,3\}$ can be expected to be the least complete. However, we expect the low-loss regions of the LFL to be sampled extensively. For $H = 1$, we also compare the LFLs for $N_{\text{data}} \in \{100,\ 250,\ 500,\ 1000,\ 2000,\ 3000,\ 4000,\ 5000,\ 10\,000,\ 20\,000,\ 30\,000,\ 40\,000, 50\,000, 60\,000, 70\,000, 80\,000, 90\,000\}$ to confirm these observed trends, for which Tab. 3.2 holds the corresponding data[2].

[2] It is worth noting that n_{min} and n_{ts} show clear monotonic trends only when N_{data} is moderately small and when n_{min} is sufficiently large. For $N_{\text{data}} > 40\,000$, the small numbers of n_{min} and n_{ts} begin to oscillate. We attribute this to insufficient convergence, which is likely for cases with large N_{data}, as the computational cost scales linearly with N_{data} and is thus very high in these cases. It could be possible that the few remaining minima above the global minimum in fact feature extremely small downhill barriers, which can vanish as a consequence of small fluctuations. This organisation could result in a disappearance and reappearance of these minima in the LFL as different data subsets of the full dataset are selected for training. However, this statement is speculative and would require confirmation through more careful analysis. We also note that n_{ts} is surprisingly large for $N_{\text{data}} = 10\,000$, which we expect arises because a more thorough TS search is performed for this LFL. This observation again suggests the incompleteness of the minimum and TS databases presented here for large N_{data}.

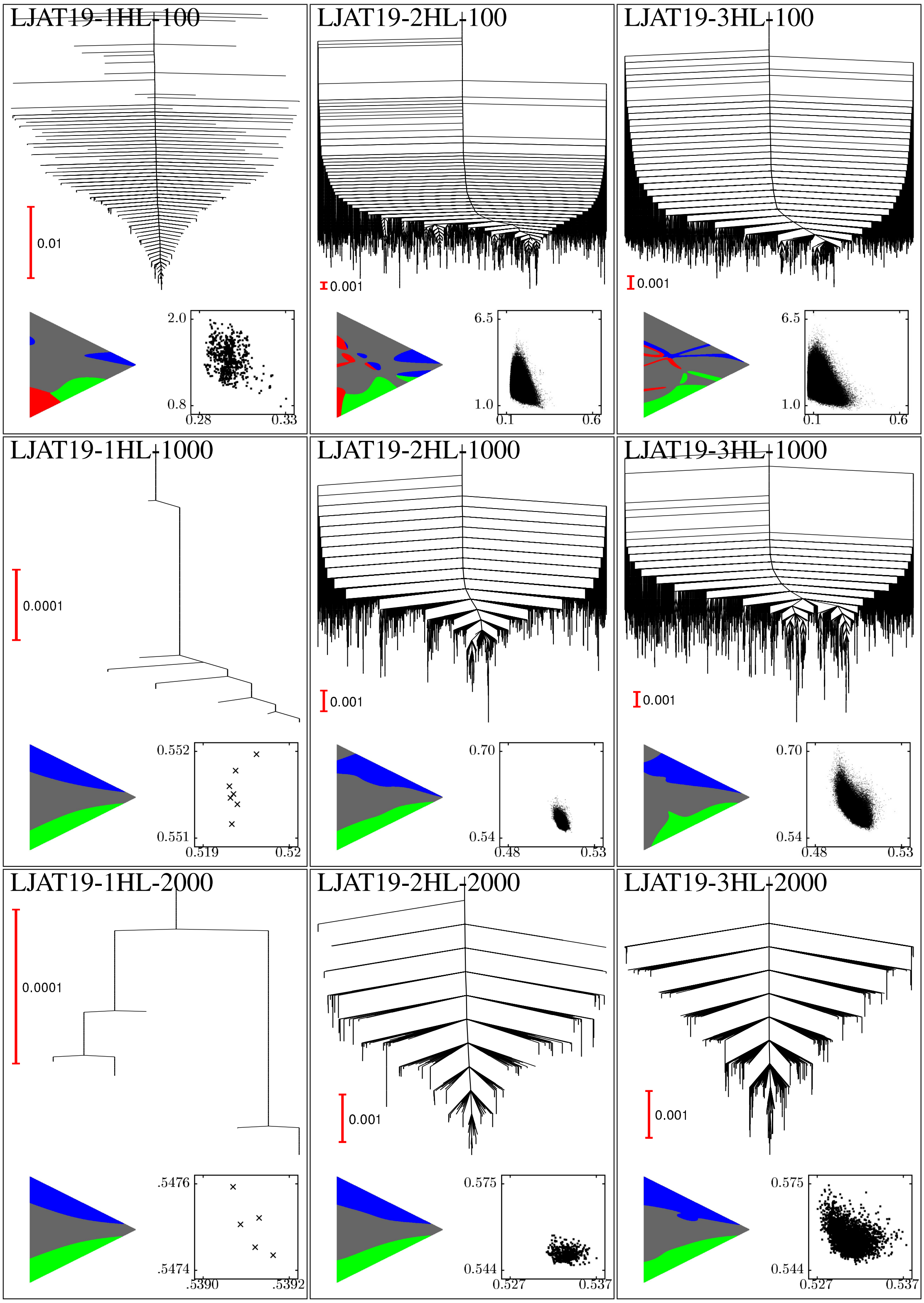
Table 3.2 Number of local minima and TSs discovered in the LFL of a DNN with $H = 1$ as a function of N_{data} using the LJAT19 dataset.

N_{data}	100	250	500	1000	2000	3000	4000	5000	10 000
n_{min}	649	436	37	7	5	4	8	2	7
n_{ts}	10 426	5867	244	33	21	18	21	10	43

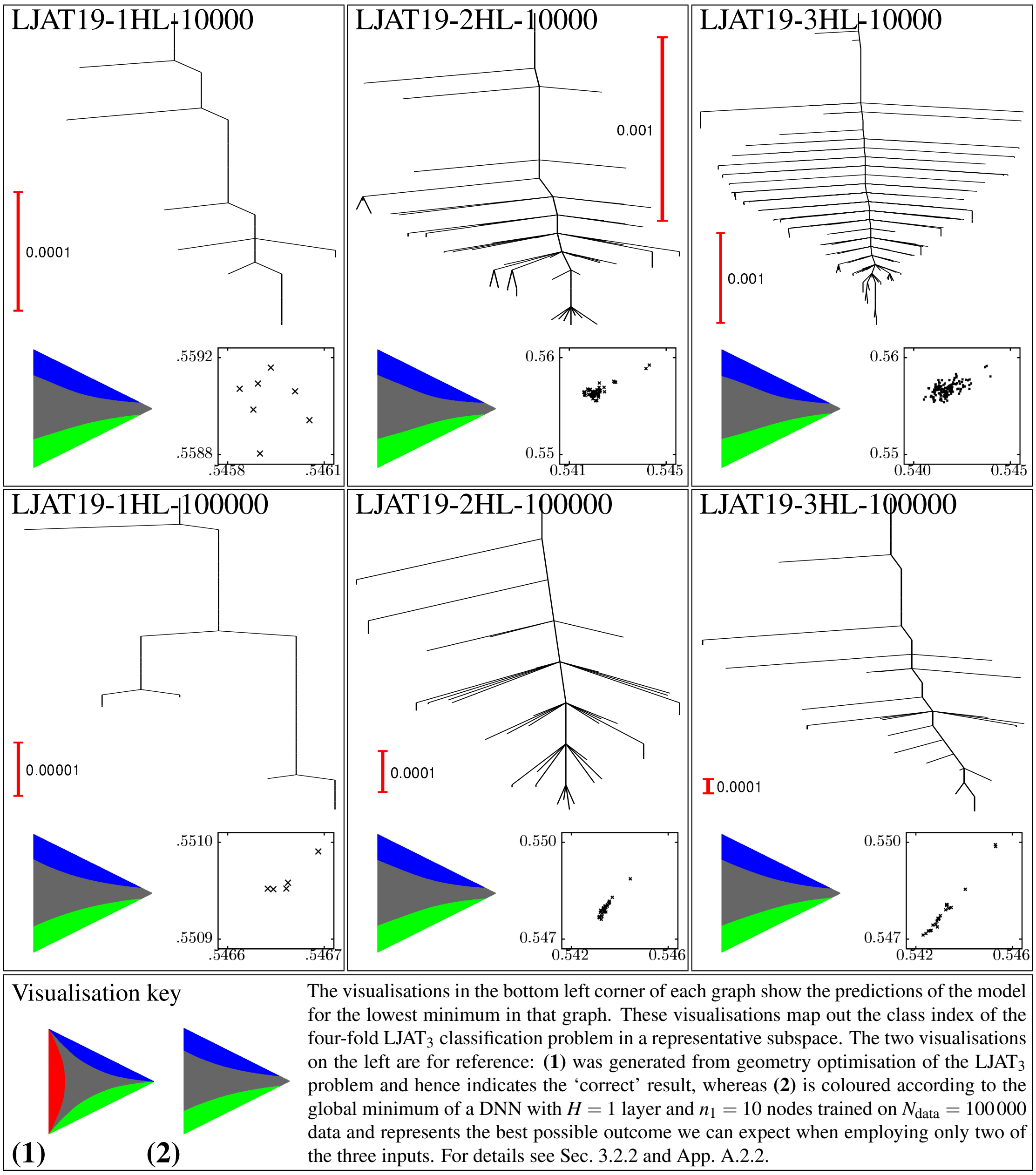
N_{data}	20 000	30 000	40 000	50 000	60 000	70 000	80 000	90 000	100 000
n_{min}	6	7	3	6	3	3	8	4	5
n_{ts}	19	4	5	3	7	8	11	6	11

A more interesting observation that can be made analysing the data reported in Tab. 3.1 is that n_{min} and n_{ts} for fixed N_{data} increase drastically when going from the shallow limit with $H = 1$ to the DNN case with $H = 2$. We stress that this is seen despite the fact that the number of variables is kept approximately fixed between those DNN architectures. This observation can be interpreted as an increased complexity of the LFL for deeper compared to shallower DNNs. A similar increase from $H = 2$ to $H = 3$ is observable for almost all values of N_{data} too, albeit to a much lesser extent. It should also be noted that due to the large number of existing minima and TSs for $H \geq 2$ and due to the difficulty with identifying permutational isomers further explained in App. A.3, a full convergence of the numbers shown in Tab. 3.1 cannot be guaranteed, and might lead to an underestimation or overestimation of the change from $H = 2$ to $H = 3$.

As is further outlined in Sec. 3.2.6, a common goal in ML is to identify the best-performing ML model among a range of options. The best-performing model is characterised by high expressibility of the fitting function accompanied by fewest possible optimisable parameters (since the training and evaluation cost increases with this number). Comparisons between DNN architectures are therefore often made between models of the same dimensionality of the parameter space, ν. In fact, the success of deep learning is due to the increased expressibility of the fitting function of deep compared to shallow DNNs and the fact that training deeper networks remains feasible. Our present observation of the strong increase in number of minima for deeper networks at fixed ν allows us to draw the interesting conclusion that the strong increase in the number of local minima does not seem to have a great impact on optimisation performance, which strengthens our motivation to investigate the connectivity between minima. We shall revisit the interpretation of this finding at the end of this chapter.



(This figure is continued on the next page.)



The visualisations in the bottom left corner of each graph show the predictions of the model for the lowest minimum in that graph. These visualisations map out the class index of the four-fold LJAT$_3$ classification problem in a representative subspace. The two visualisations on the left are for reference: **(1)** was generated from geometry optimisation of the LJAT$_3$ problem and hence indicates the 'correct' result, whereas **(2)** is coloured according to the global minimum of a DNN with $H = 1$ layer and $n_1 = 10$ nodes trained on $N_{\text{data}} = 100\,000$ data and represents the best possible outcome we can expect when employing only two of the three inputs. For details see Sec. 3.2.2 and App. A.2.2.

Reprinted with permission from Ref. [142], P.C. Verpoort et al., Proc. National Academy of Sci. USA 117, 21857 (2020). © *2020 National Academy of Sciences.*

Fig. 3.2 Disconnectivity graphs for $N_{\text{data}} \in \{100, 1000, 2000, 10\,000, 100\,000\}$ training data (from top to bottom) for the DNNs with $H \in \{1,2,3\}$ hidden layers (from left to right), labelled as 'DATASET-#HL-#' on the top of each panel, where the two digits indicate, respectively, the number of hidden layers, H, and the amount of training data, N_{data}. Only the lowest 2000 minima (or all if fewer than 2000 were identified) are shown, and the vertical scale has been adjusted to span the range of loss-function values within this set. Included as an inset graph below each disconnectivity graph are a graphical visualisation of the performance of the global minimum (see App. A.2 for details) as well as a plot of the training (horizontal axis) versus testing (vertical axis) loss values of all minima. It is apparent from these graphs that in each case the structure of the LFL is either funnelled or comprises many minima with similar loss values connected by low barriers.

3.2.2 Connectivity of minima

Next, we report the disconnectivity graphs for LFLs using the LJAT19 dataset, which can be found in Fig. 3.2. Note that these disconnectivity graphs show only the lowest 2000 minima for reasons of clarity. It is evident from those graphs that for sufficiently large N_{data} the loss-function features a relatively simple structure corresponding to a funnelled landscape, where finding the global minimum[3] is straightforward [145]. This 'palm tree' [150] organisation becomes particularly clear in panels[4] LJAT19-1HL-100, LJAT19-2HL-2000, LJAT19-3HL-10000, and those with higher N_{data}: pathways from all the minima with higher loss to the global minimum encounter relatively low downhill barriers. The structure in the cases with small N_{data} and $H \in \{2,3\}$ is rather different: we see many low-lying minima with similar loss values. Moreover, it can be seen that the barriers between the minima are very low, especially those for LJAT19-2HL-100 and LJAT19-3HL-100, where the number of minima is the highest, which is in agreement with previous calculations [45]. This should be contrasted with the PELs that have previously been reported for structural glasses [39–41, 102, 103]. The LFLs of these systems equally feature a large number of amorphous structures with similar potential energy. In contrast, however, these minima are separated by large barriers compared to the relevant thermal energy, producing a landscape with a hierarchical structure [150]. The structure at the bottom of the LFLs for cases LJAT19-2HL-100 and LJAT19-3HL-100 has previously been unobserved, and the disconnectivity graph is reminiscent of a 'mangrove swamp'. The overall funnelled structure could serve as an explanation for the unexpected efficiency of commonly employed DNN optimisation methods, as is discussed further at the end of this chapter.

Visualisations of the predictions of the DNN global minimum are included in Fig. 3.2; a key is provided below the figure with further details in App. A.2.2. These visualisations plot the LJAT19 class index in a representative subspace of the full three-dimensional space of LJAT19 input features. This is achieved through projecting the three input features onto a plane, where the details of this procedure are outlined in App. A.2.2. The four-fold class index is plotted as colours grey, red, green, and blue. For further details, we refer the reader to the appendix, but we stress again that these visualisations can simply be regarded as one way of displaying the classification outcomes of DNN models in a reduced-dimensionality subspace of the full three-dimensional space of input features. These visualisations allow

[3] Again we remind the reader that our main interest in the context of ML lies with the optimisability of the LFL structure and not reaching the global minimum. Identifying points in the LFL with low training loss is sufficient, and these points need not be the lowest minimum (and in fact this is even counterproductive, as further discussed in Sec. 3.2.3).

[4] We remind the reader that the 'DATASET-#HL-#' notation used to refer to a specific LFL example is introduced at the beginning of Sec. 3.2.

us to observe how the best solution converges to the same pattern as sufficient training data are supplied (see Fig. 3.2 as N_{data} increases from top to bottom). The best AUC value here is around 0.8 and corresponds to convergence of the relative probabilities for predictions in the subspace of one missing input variable. Note that the class index coloured in red is entirely absent in these visualisations, reflecting the difficulty of distinguishing the red from the grey class index in the LJAT$_3$ problem without knowledge of all three inputs. The same optimal solution is obtained for $H \in \{1,2,3\}$, but more training data are required for larger H (even though the number of variable weights that are optimised is very similar in each case). The enhanced expressibility of the DNNs with higher H may be reflected in the increased complexity of the patterns in the visualisations (see especially $N_{\text{data}} \in \{100, 1000\}$).

The performance of all minima in the LFLs for the LJAT19 training dataset can also be analysed for the accompanying testing dataset. We plot the training versus testing loss of all minima as scatter plots in Fig. 3.2. Because n_{min} is large in the cases in which $H \in \{2,3\}$ and $N_{\text{data}} \in \{100, 1000, 2000\}$ and because the resulting high density of points in the scatter plots presented in Fig. 3.2 makes these plots unreadable, we additionally provide histograms in Fig. 3.3 for the relevant cases (i.e. $H \in \{2,3\}$ and $N_{\text{data}} \in \{100, 1000, 2000\}$). The graphs for $H \in \{2,3\}$ reveal the following trends: there exists a weak anti-correlation between the train-loss and test-loss values of low-lying minima in the case of little training data ($N_{\text{data}} \leq 1000$). This result suggests overfitting: minima with loss values lower than the optimal value obtained for the large training data limit must gain their advantage in a way that does not generalise well to testing data. In this regime, the corresponding DNN models likely yield no good learning capability when trained. This result is due to the lack of train-test loss correlation, which results in poor generalisability of the model. For a medium amount of training data ($N_{\text{data}} = 2000$), the graphs indicate no clear correlation between the relative variations of train and test loss, while for large amounts of training data ($N_{\text{data}} \geq 10\,000$) there is a positive correlation for, both, absolute and relative loss variations. Interestingly, the lack of bare train-test loss correlation for medium amounts of training data, e.g. $N_{\text{data}} = 2000$ (but not small, e.g. $N_{\text{data}} = 100$) can be overcome by considering the correlation with the minimum geometry, as discussed in the next section.

3.2.3 Relating test error and basin geometry

To further investigate the correlation of training and testing loss, L_{train} and L_{test} as well as how the loss of each minimum correlates with local curvatures, we perform a fit of the following

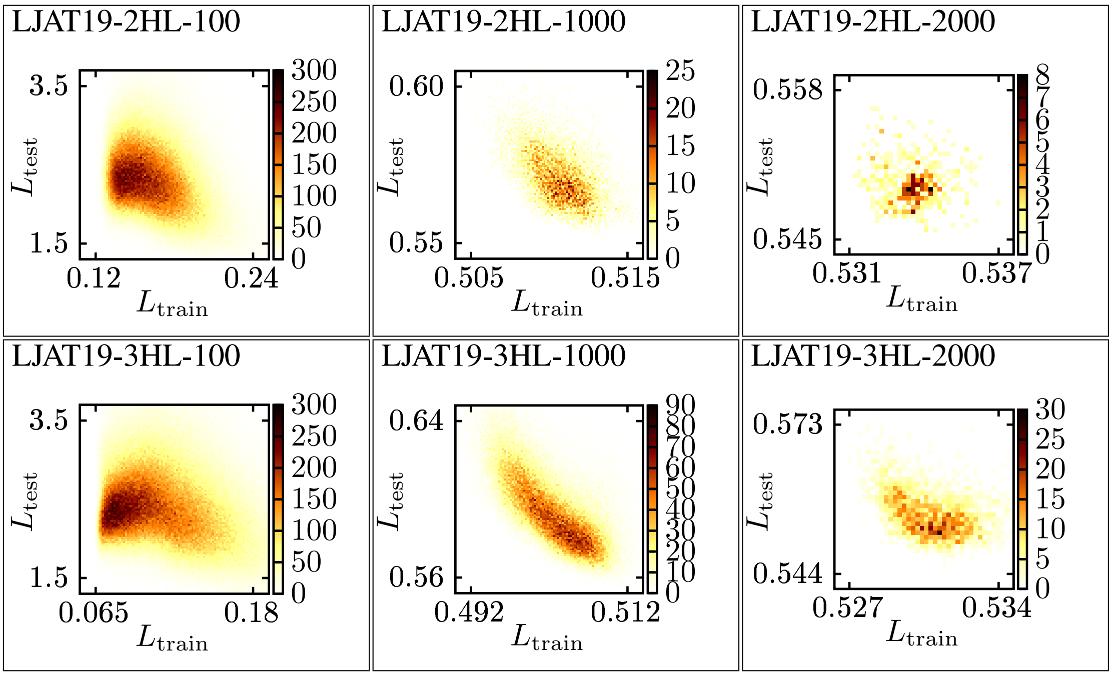


Fig. 3.3 Minima of the LFLs for the LJAT19 training dataset with $H \in \{2,3\}$ and $N_{\mathrm{data}} \in \{100, 1000, 2000\}$ presented as histograms as a function of training and testing loss, L_{train} and L_{test}. The number of minima in one bin of the histogram is indicated by its colour, and the colour can be interpreted using the colour scale on the right-hand side of each histogram.

two functions to the data:

$$L^{(1)}_{\mathrm{test}}(L_{\mathrm{train}}) \quad = a_1 + b_1 L_{\mathrm{train}} \tag{3.1}$$

and

$$L^{(2)}_{\mathrm{test}}(L_{\mathrm{train}}, S) = a_2 + b_2 L_{\mathrm{train}} + c_2 S, \tag{3.2}$$

where S is the log product of all eigenvalues of the Hessian matrix evaluated for the respective minima and is defined analogously to the entropy in molecular systems. The optimal fit parameters and values of the adjusted coefficient of determination, r^2, are reported in Tab. 3.3, where fit results for $H = 1$ and $N_{\mathrm{data}} \geq 1000$ were omitted (because $n_{\mathrm{min}} \leq 7$, which does not allow a conclusive statistical analysis). First, the trend of negative to positive correlation between L_{train} and L_{test} as a function of increasing N_{data} is confirmed by the values of b_1 and b_2. Second, while adding the term proportional to S seems to be irrelevant in the case of $N_{\mathrm{data}} \geq 10\,000$ (as c_2 is small, b_2 is similar to b_1, and r^2 changes only slightly), the results are very different for small N_{data}: the parameter c_2 increases by up to three orders of magnitude

Table 3.3 Fitting parameters for correlation analysis between training and testing loss values of minima (see Eqs. 3.1 and 3.2).

N_{data}	b_1	r^2	b_2	c_2	r^2
$H = 1$					
100	−7.8(11)	0.068	−4.7(11)	8.39(88)	0.18
$H = 2$					
100	−5.500(27)	0.097	−4.858(27)	5.57(5)	0.12
1000	−2.529(38)	0.25	−0.308(33)	0.4495(38)	0.63
2000	−0.155(85)	0.0047	0.231(70)	0.1510(85)	0.4
10000	1.07(10)	0.68	0.90(13)	0.035(16)	0.71
100000	0.940(57)	0.89	0.944(58)	0.0034(47)	0.89
$H = 3$					
100	−4.050(17)	0.09	−4.678(16)	6.736(30)	0.16
1000	−2.998(11)	0.48	−1.083(13)	0.5131(25)	0.65
2000	−1.059(45)	0.15	−0.063(35)	0.2122(37)	0.6
10000	0.673(75)	0.35	0.672(71)	0.0267(69)	0.41
100000	0.945(23)	0.99	0.939(25)	−0.0019(27)	0.99

with decreasing N_{data}, the optimal values of b_1 and b_2 differ drastically, and the r^2 value increases significantly between the two fits. The change between fitting with Eq. (3.1) and Eq. (3.2) is most pronounced for $N_{\mathrm{data}} \in \{1000, 2000\}$. We even find that in the case of LJAT19-2HL-2000 the correlation between training and testing loss changes from negative to positive, and r^2 increases from 0.0067 to 0.4. This result suggests that knowledge of the curvature of a minimum (in our case encoded in the entropy parameter S) may enhance the prediction of the performance of training LFL minima on testing data, which is in good agreement with studies by Wu et al. reported in Ref. [155]. We note that S summarises the information contained in the eigenvalues of all the ν dimensions of the Hessian matrix (where we recall that in the present example ν takes the values 74, 69, and 72 for 1, 2, and 3 hidden layers respectively). In future work, it would be interesting to investigate the distribution of those eigenvalues [117], which play a key role in determining the structure of the heat capacity analogue for the LFL [12, 36].

3.2.4 Performance of minima

We now turn to the analysis of the performance of LFL minima for both training and testing datasets. In Tab. 3.4, we list the loss function and AUC values averaged over all minima for both datasets. The values reported in that table display the following trends: as N_{data} increases, the average performance improves. Training loss increases, training AUC decreases, testing

Table 3.4 Average loss function and AUC values for train and test data of all minima found in training as listed in Tab. 3.1.

	$H = 1$				$H = 2$			
	train		test		train		test	
N_{data}	loss	AUC	loss	AUC	loss	AUC	loss	AUC
100	0.297	0.953	1.453	0.547	0.166	0.995	2.344	0.557
1000	0.519	0.810	0.552	0.796	0.511	0.823	0.572	0.787
2000	0.539	0.806	0.548	0.795	0.534	0.809	0.550	0.792
10000	0.546	0.801	0.559	0.801	0.542	0.802	0.557	0.801
100000	0.547	0.797	0.551	0.796	0.543	0.798	0.548	0.797

	$H = 3$			
	train		test	
N_{data}	loss	AUC	loss	AUC
100	0.121	0.994	2.424	0.564
1000	0.502	0.829	0.596	0.779
2000	0.531	0.812	0.556	0.788
10000	0.542	0.803	0.557	0.801
100000	0.543	0.798	0.548	0.797

loss decreases, and testing AUC increases with growing N_{data}. We note that training loss *should* decrease with growing N_{data}, as this indicates a reduced overfitting of the data. Weak deviations from this behaviour persist when the model is already converged, and the numbers only change by less than 1 %, which we attribute to statistical noise. We note that this effect could be linked to the variations of n_{min} and n_{ts} reported in Sec. 3.2.1.

Next, we test the average model performance as a function of minima included in the assessment, with results for $H \in \{2,3\}$ and $N_{\text{data}} \in \{100, 1000, 2000\}$ shown in Fig. 3.4. Interestingly, for all cases except for the ones with fewest training data provided (i.e. $N_{\text{data}} = 100$), the average minimum performance on test data is worse at the bottom of the training loss landscape compared to the overall average performance. Note that the graphs show the reduced loss function, which reports loss values ranging from 0 to 1, where 0 would correspond to the lowest minimum and 1 to the highest discovered in the respective LFL. The absolute loss values can therefore not be compared between the graphs, yet they show how the loss values of minima relative to all minima of the respective landscape perform. The trends revealed by these graphs are intriguing: the lowest minima of the $N_{\text{data}} = 100$ landscape in average perform better (compared to all other minima in that specific LFL) than in the case for $N_{\text{data}} \geq 1000$. However, note that these results reported in Fig. 3.4 may be particularly susceptible to the undersampling of minima in loss ranges of high minima density discussed at the beginning of Sec. 3.2 and in App. A.3.

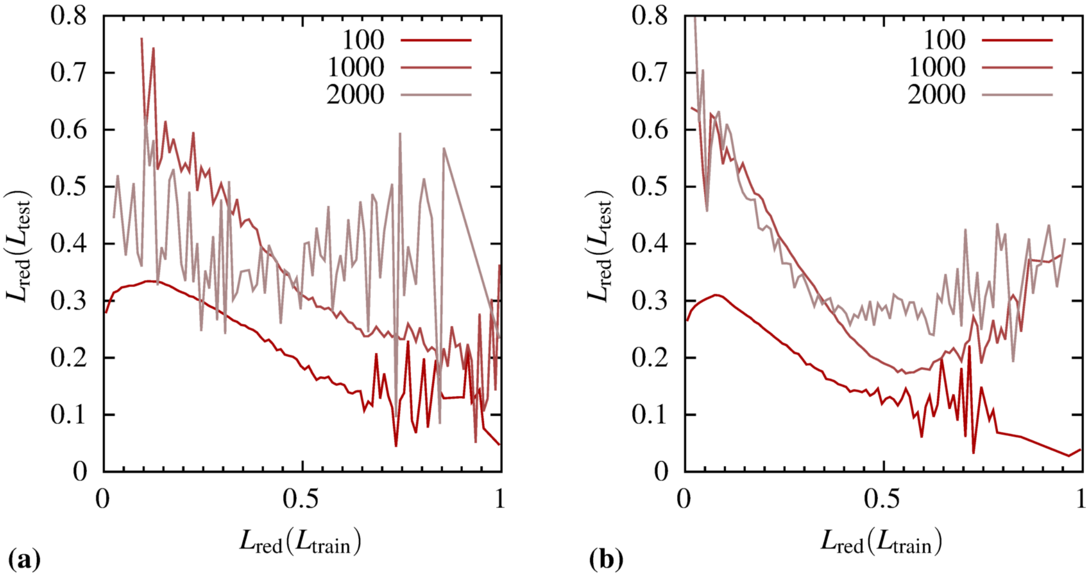


Fig. 3.4 Reduced test loss plotted against reduced train loss of minima of the LFL for the LJAT19 dataset with $N_{\text{data}} \in \{100, 1000, 2000\}$. The train loss is divided into 100 intervals and the test loss is averaged over all minima found with train loss in the interval. The reduced loss is defined as $L_{\text{red}}(L) = \frac{L - L_{\text{min}}}{L_{\text{max}} - L_{\text{min}}}$, where L_{max} is the maximal and L_{min} is the minimal loss value in the corresponding database of minima. The graph shows that the average test loss increases towards the bottom of the train loss landscape for $N_{\text{data}} \in \{1000, 2000\}$, as one would intuitively expect from overfitting. Intriguingly, for $N_{\text{data}} = 100$, the average test loss seems to decrease again at the bottom of the training loss landscape. This result indicates that, when very little N_{data} is used, minima very close to the bottom of the landscape overfit slightly less than ones further up in the landscape. **(a)** $H = 2$. **(b)** $H = 3$.

3.2.5 Barrier heights

In Tab. 3.5, we report the barrier heights to the global minimum averaged over all minima discovered in the LFL, together with the average of those barrier heights divided by the loss difference. Intriguingly, the relative barrier heights are small in all cases, and especially the cases with $H \in \{2, 3\}$ and $N_{\text{data}} \in \{100, 1000\}$, where the disconnectivity graphs in Sec. 3.2.2 reveal a ‘mangrove-swamp’ type structure with many local minima of similar loss values, are characterised by extremely low relative barriers (defined as the barriers heights divided by the loss difference between the minima). These relative barrier heights are below 10 %, as is evident from the numbers reported in Tab. 3.5.

It is also interesting to discuss trends in the barrier heights across different values of H and N_{data}. It is clear that the total barrier heights decrease monotonically in all those cases in which n_{min} is of reasonable size (i.e. not for $H = 1$ with $N_{\text{data}} \geq 2000$). Clear trends in the

Table 3.5 Downhill barrier of a minimum to the global minimum for all other minima located in the training LFL, averaged over all minima. The ‘relative’ column reports the average of these barrier heights divided by the loss difference between the minima.

	$H = 1$		$H = 2$		$H = 3$	
N_{data}	total	relative	total	relative	total	relative
100	0.105×10^{-3}	0.0178	0.276×10^{-2}	0.0783	0.324×10^{-2}	0.0781
1000	0.298×10^{-5}	0.0905	0.357×10^{-3}	0.0562	0.540×10^{-3}	0.0462
2000	0.589×10^{-4}	0.9777	0.101×10^{-3}	0.0676	0.566×10^{-4}	0.0294
10000	0.415×10^{-5}	0.0597	0.316×10^{-4}	0.4342	0.324×10^{-4}	0.0477
100000	0.663×10^{-5}	0.3301	0.332×10^{-4}	0.3019	0.216×10^{-4}	0.0286

relative barrier heights would be even more interesting but can unfortunately not be identified in the data presented in Tab. 3.5. We note that the average barrier heights are particularly susceptible to full convergence of minima and TS databases, which could be the reason for the absence of trends in the relative barrier heights in the reported data. We conclude that the main result obtained from the presented data is that the barriers are all low, whereas we are unable to deduce any clear trends as a function of H or N_{data} from them.

These results are of great relevance to the understanding of the effectiveness of SGD and similar methods for the optimisation of LFLs of DNNs. The typical barriers that an optimiser has to overcome are small compared to the overall change of the loss function. This structure means that small amounts of noise added to the loss value and its gradient are sufficient to result in a global downhill trajectory even for a ‘simple’ gradient-based optimiser.

3.2.6 Performance as a function of number of hidden nodes

The main results reported in this chapter focus on systems with varying deepness of DNNs while keeping the number of weight variables, v, approximately fixed. This setup is dominantly used as it allows a direct comparison of the obtained results in previous work on shallow neural networks [11, 12, 30, 36, 37]. Before we move on to present results that do not obey this constraint, we give a list of other reasons for this choice of comparison.

First, it is generally known for PELs that the complexity of a landscape increases with v, and so do the numbers of minima and TSs, n_{min} and n_{ts}. In fact they do so exponentially with the number of degrees of freedom D as $n_{\text{min}} \propto \exp(\gamma D)$ and $n_{\text{ts}} \propto D\exp(\gamma D)$ [133, 149]. We would therefore like to find out what happens to the structure of the LFL when the number of degrees of freedom is kept fixed, and only the deepness of the network is varied.

Second, recent studies of DNN learning have already established that the capability of DNNs to some extent stems from over-parameterisation [9]. While this result appears to

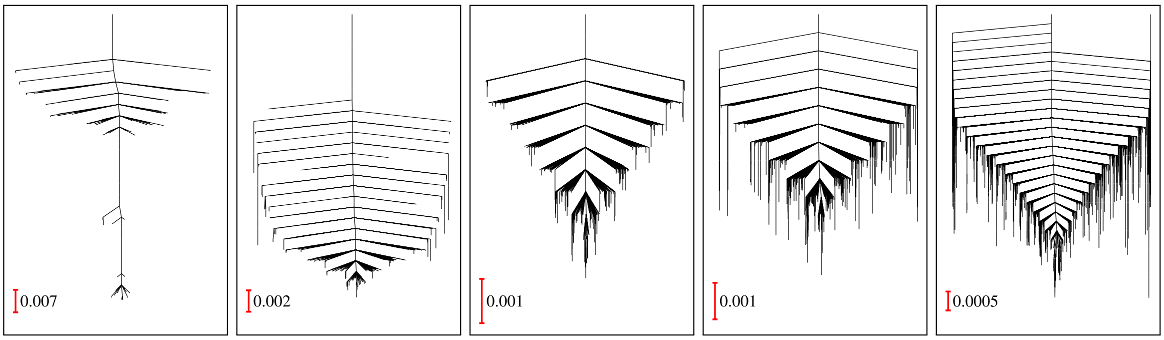


Fig. 3.5 Disconnectivity graphs for LJAT19-3HL-2000 with $n_l \in \{2,3,4,5,6\}$ nodes in each hidden layer (with n_l increasing in the order of appearance from left to right).

be a necessary condition for good performance, it certainly cannot be a sufficient condition as wide shallow neural networks are not as effective as ones that are deep but narrow. It is therefore interesting to investigate not just the degree to which over-parameterisation facilitates learning but instead how the distribution of these degrees of freedom across the network affects performance.

Third, the field of ML seeks to find the best-performing architecture in a competition between networks of similar degrees of freedom [84, 109]. This objective is due to the fact that the training cost grows at least linearly with the number of parameters, as does the cost for the evaluation of the network function in prediction.

Nonetheless, for reasons of completeness, disconnectivity graphs for the LJAT19 dataset for $N_{\text{data}} = 2000$, $H = 3$, and $n_1 = n_2 = n_3 \in \{2,3,4,5,6\}$ are presented in Fig. 3.5. As expected, n_{min} and n_{ts} grow significantly with ν. The structure of the LFL however does not vary significantly between different numbers of hidden nodes, although the landscape becomes more funnelled with a decrease of hidden nodes. For a larger number of hidden nodes, it is closer to the 'mangrove-swamp' structure, with many local minima of similar loss values connected by low barriers (which is also observed in Sec. 3.2.2 for very little training data).

3.2.7 Comparison with other datasets

To check that our results are transferable to very different prediction problems, we analyse the LFL for the OPTDIG and WINE datasets and present their disconnectivity graphs in Fig. 3.6. The two DNN architectures employed in conjunction with the OPTDIG dataset are characterised by $H = 1$, $n_l = 5$ and $H = 3$, $n_l = 4$ with, in both cases, $n_0 = 64$, $n_{H+1} = 10$. The resulting numbers of weight variables are $\nu = 385$ and $\nu = 350$, respectively. Similarly, the two architectures used for the WINE dataset are characterised by $H = 1$, $n_l = 5$ and

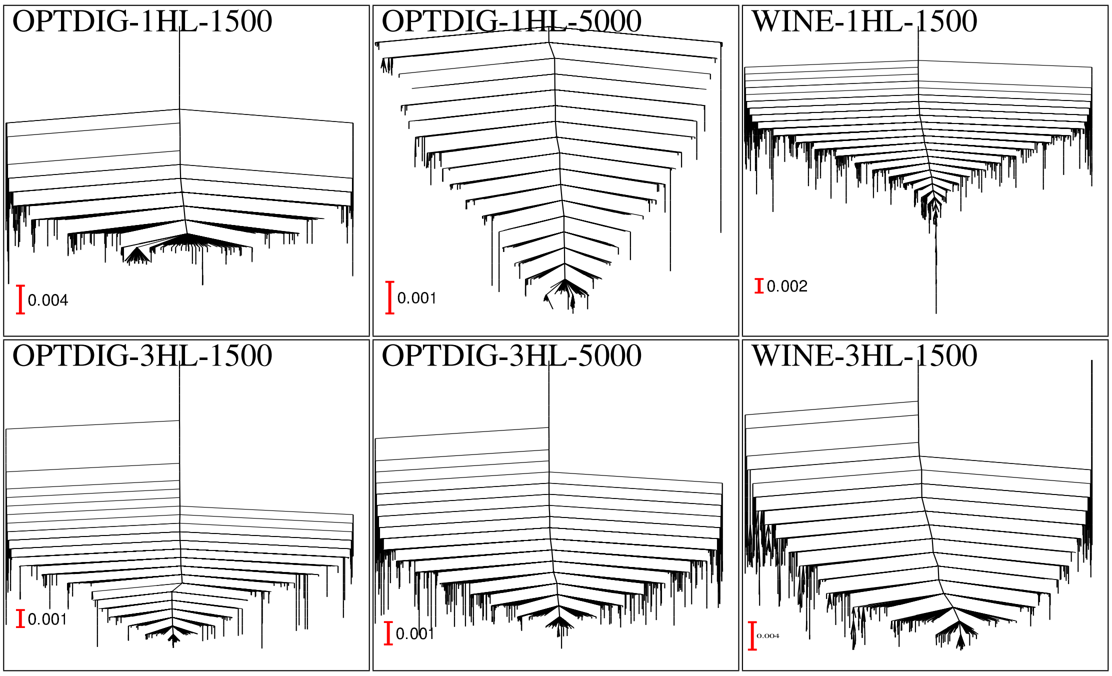
$H = 3$, $n_l = 3$ with, in both cases, $n_0 = 11$, $n_{H+1} = 3$. This results in $\nu = 78$ and $\nu = 72$. We note that for the cases studied in this section, the procedure outlined in Sec. 3.1.2 for searching minima and TSs is not converged fully, as we only aim to obtain enough minima and TSs in order to resolve the high-level LFL structure and plot the resulting disconnectivity graphs.

The structure of these landscapes appears to be qualitatively similar to those using the LJAT19 dataset. The LFLs all display a single-funnelled structure, as is evident from the disconnectivity graphs plotted in Fig. 3.6. Importantly, there are no large barriers separating low-lying local minima, which would hinder efficient optimisation. Notably the differences between the disconnectivity graphs for the different cases studied in this section is not very large. This is unsurprising given the range of N_{data} considered is much smaller than in those previously studied (N_{data} only differs by a factor of 3.3, whereas it changes over three orders of magnitude in the cases reported in Fig. 3.2). We also observe very little change in the LFL structure between $H = 1$ and $H = 3$. This may be due to insufficient convergence of the minima and TSs databases. It can be expected that the LFLs with $H = 3$ will develop a mangrove-swamp like structure as the databases are further converged. We also note that the value of N_{data} where the different LFL archetypes would emerge will depend on the precise nature of the training data (features, noise, correlation, etc). In summary, there is great scope for more studies of LFLs using other training datasets, and the results presented in this section could be developed further. Yet the key result that we are able to demonstrate here is that, again, we encounter predominantly low barriers separating local minima, which would again suggest good optimisability of the respective DNN models.

3.3 Conclusions

We conclude this chapter by summarising and interpreting the results reported above. The central study we present is the resolution of the structure of the LFL of DNNs as a function of amount of training data, N_{data}, provided and for varying deepness, as captured by the number of hidden layers, H. This structure is analysed in a number of ways, of which the main findings are again repeated here.

In Sec. 3.2.1, we report the number of minima, n_{min}, and the number of TSs, n_{ts}, which both decrease with growing N_{data}, as expected. This result is in agreement with the expected exponential growth with the number of degrees of freedom D as $n_{\text{min}} \propto \exp(\gamma D)$ and $n_{\text{ts}} \propto D\exp(\gamma D)$ found in PELs [133, 149]. More interestingly, both n_{min} and n_{ts} increase significantly as H is increased from 1 to 2, hence indicating that the ‘deep’ networks behave rather differently from the ‘shallow’ ones.



Reprinted with permission from Ref. [142], P.C. Verpoort et al., Proc. National Academy of Sci. USA 117, 21857 (2020). © *2020 National Academy of Sciences.*

Fig. 3.6 Disconnectivity graphs for the training datasets OPTDIG (with $N_{\text{data}} \in \{1500, 5000\}$ and $H \in \{1,3\}$) and WINE (with $N_{\text{data}} = 1500$ and $H \in \{1,3\}$). Only the lowest 2000 minima (or all of them if fewer than 2000 were found) are shown. The vertical scale is adjusted to span the range of loss-function values within this set.

This trend is substantiated in Sec. 3.2.2, where we plot the disconnectivity graphs. The graphs reveal one of the main findings of our study: the LFLs of DNNs fall into two archetypes depending on N_{data} and H. We either observe a single-funnelled landscape, when N_{data} is large or when $H = 1$. For the data-scarce deep examples, however, a new, previously unreported structure is discovered with many minima of similar loss values connected by small barriers, which we call a 'mangrove-swamp' type structure due to the appearance of the disconnectivity graph. This result also shows that the LFL of a DNN is qualitatively different from the PEL of a structural glasses [39–41, 102, 103] or systems with multi-funnelled landscapes [27].

Next, in Sec. 3.2.3, a study of the correlation between train loss, test loss, and minimum curvature (quantified using the log product of Hessian eigenvalues) for all discovered minima allows us to recover results previously reported [155] that attribute the success of SGD to the wide basin geometry of 'good' minima. When little training data is supplied, the large number of minima in the LFL for the training dataset spread across a wide range of loss values, and the same is true for loss values of those minima when evaluated for the accompanying testing dataset. In this limit, the system is prone to overfitting, which manifests as an anticorrelation between training and testing loss values. Moreover, a strong correlation between the local curvatures and the generalisability of minima in the deep medium-N_{data} regime is observed. Hence we conclude that entropic contributions to the optimisation dynamics may guide SGD-type optimisers to minima that generalise well. This observation allows us to gain deeper insight into the success reported for additional mechanisms that guide the optimiser towards low-curvature minima [10, 26].

Finally, an analysis of the barrier heights in Sec. 3.2.5 confirms the assessment from Sec. 3.2.2 that the barriers disconnecting local minima are small, especially when compared to the respective loss value difference, and this observation holds for all network architectures irrespective of N_{data}.

The main conclusion drawn from these results is that our analysis of the LFL structure helps in answering the question why simple optimisation procedures such as SGD are successful for systems as complex as DNNs with a high-dimensional weight-variable space and non-convex loss function. The LFL is either funnelled or exhibits a structure with many competing low-loss minima connected by low barriers. Both structures are easy to optimise as the barriers that any optimiser has to overcome in order to follow a global downhill trajectory are small compared to the overall loss value changes.

We stress that this is the first study of its kind to be conducted. Previous works either only considered single-hidden layer neural networks [11, 12, 30, 36, 37], employed nudged-elastic band approaches but without properly assessing the connections [45], or inferred properties of the LFL from the training dynamics of DNNs [9]. While these previous contributions are relevant and add other interesting observations not studied in this work, our in-depth analysis of the LFL provides a much more complete picture.

We reiterate that the methods employed in this work are in principle capable of discovering all minima and TSs in the LFL of DNNs. However, the problem of the potential miss-identification of stationary points as permutational isomers due to a high density (as a function of training loss) of such stationary points remains a challenge that this work is not able to fully resolve. Future work will have to find a more suitable approach for dealing with this problem in order to discover sufficiently converged databases of minima and TSs.

It would be interesting in future work to extend the approach presented here to larger DNN architectures that are closer to the ones commonly used in practical applications. The trends for small architectures that we report in this contribution appear to be consistent with analytical studies based on spin-glass models [31]. Yet, developing the numerical methods used in this contribution to the point where they are efficient enough to resolve at least moderately large structures is desirable. However, this aim is unfortunately not straightforward due to the significant growth in computational cost. The numerical results obtained and presented in this work required an approximate 170000 core hours, which can be expected to grow strongly with increasing complexity of network architecture and size of training dataset. First, the methods that are outlined in Sec. 3.1.2 and that are employed to obtain the results reported throughout this chapter scale at least linearly with the dimensionality of the space, which in our case is the number of weight variables, ν. Second, sampling the entire space becomes increasingly difficult because the number of minima, n_{min}, and number of TSs, n_{ts}, grow quickly, which is known to be exponential for PELs [133, 149], and has been demonstrated to be of equal severity for LFLs in Sec. 3.2.1. It is possible that sampling methods that identify small subsets of minima and TSs for detailed analysis, dimensionality-reduction techniques that eliminate less relevant directions (either particularly 'soft' or particularly 'hard' ones) or an increased usage of the regularisation to reduce the effective n_{min} and n_{ts} could be viable approaches to overcome the impediments arising from increasing computational cost in larger systems.

Many open research questions regarding the structure of the LFL and its optimisation with SGD remain, of which several can be addressed using the methods presented in this chapter. For example, higher-index stationary points (those with Hessian index greater than 1) would be interesting to study more carefully due to their possible relevance for deep learning optimisation, given that many DNN optimisers cannot distinguish between minima and other types of stationary points. Moreover, it would be interesting to investigate other ML architectures such as convolutional neural networks, recurrent neural networks, or DNNs with reduced node connectivity. Notably, previous studies reported exciting effects of reduced node connectivity on the LFL structure of shallow networks [30], and it would be interesting to extend those studies to deep networks.

In summary, this work has advanced our understanding of the LFLs of DNNs, yet many open tasks remain that can be addressed using the methods and procedures described.

Part II

Equilibration in strongly Rashba-coupled systems

Chapter 4

Introduction to Rashba systems and non-equilibrium solid-state physics

The second part of this thesis focuses on the study of non-equilibrium charge-carrier densities in systems with strong Rashba spin-orbit coupling. Therefore, the system studied here is very different in nature to the one studied in the first part of this thesis. Nonetheless, the two presented studies are linked by the fact that, in both cases, we are investigating the relaxation of a complex system with many degrees of freedom, albeit by the use of very different methods.

The study outlined in this and the following two chapters consists of two main points of investigation. First in Chap. 5, we attempt to understand and analyse the transport measurements of two systems that are known to have strong Rashba coupling and that display intriguing non-equilibrium effects in their magnetotransport curves. Second, these observations together with our attempts to explain and model them inspire a systematic study of chirality relaxation in Rashba systems, for which results are reported in Chap. 6. The remainder of the present chapter introduces the required theoretical background that will be used throughout the next two chapters.

4.1 Rashba systems

Rashba systems are a class of materials that have proven to be of great interest for both exploitation in technological applications as well as for academic study of intriguing phenomena in solid-state systems.

The Rashba effect is named after E.I. Rashba, who, together with Y.A. Bychkov, was the first to discover and describe it [23]. This effect exists in solid-state systems with strong spin-orbit coupling together with a broken inversion symmetry and can either manifest in three-dimensional (3D) materials that have a distorted inversion-asymmetric unit cell, such as Ferroelectric materials, or in two-dimensional (2D) systems, in which the Rashba effect arises due to the existence of an electric field perpendicular to the plane, typically realised on the interface between two materials where a 2D electron gas or edge states can be found [77].

More recently, the Rashba effect has also been realised artificially and studied in ultracold atomic gases [160]. The Rashba effect and its implications on transport effects is well-documented in a number of review articles [85, 16, 17], and we summarise some main aspects relevant to our study in this section.

The Rashba effect gives rise to a spin-orbit coupling term in the Hamiltonian of the form

$$H_{\text{Rashba}} = \alpha_{\text{R}} \boldsymbol{\sigma} \cdot (\mathbf{r}_{\text{SO}} \times \mathbf{k}) \,, \tag{4.1}$$

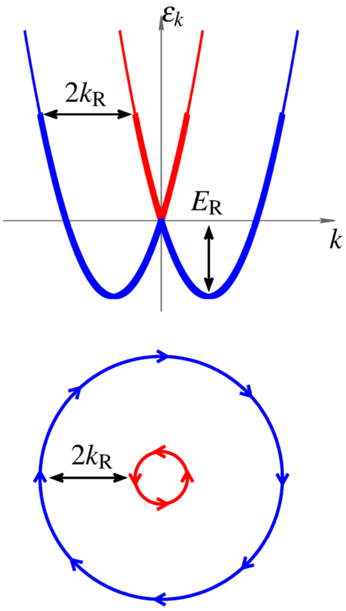
where α_{R} is the Rashba parameter, $\boldsymbol{\sigma}$ is the vector of the Pauli matrices, $\mathbf{k}$ is the wavevector, and $\mathbf{r}_{\text{SO}}$, which is a unit vector, points in the direction of the spin-orbit coupling along which inversion symmetry is broken.

For a parabolic Bloch band, i.e. $H_{\text{free}} = \hbar^2\mathbf{k}^2/2m$ with effective mass m, the additional Rashba interaction term lifts the spin degeneracy of the two Bloch bands and leads to momentum-dependent spin mixing within the band. This yields the energy dispersion

$$\varepsilon_k^{\pm} = \hbar^2/2m\,(k \pm k_{\text{R}})^2 \ - E_{\text{R}} \,, \tag{4.2}$$

where $k_{\text{R}} = \alpha_{\text{R}} m/\hbar^2$ is known as the Rashba momentum and $E_{\text{R}} = \hbar^2 k_{\text{R}}^2/2m$ is the Rashba energy scale, and where the $+$ $(-)$ superscript refers to the upper (lower) Rashba band, respectively. The energy dispersions for a Rashba system that result from Eq. (4.2) are plotted in Fig. 4.1 along with an indication of the Rashba energy scale E_{R} and the Rashba momentum k_{R}. The band index replaces the degenerate spin index of the free parabolic energy dispersion. As a consequence, the new dispersion features k-dependent spin alignments on the Fermi surface. The free-electron parabolic dispersions of the two spin species are mutually shifted apart in momentum space by the Rashba momentum k_{R}. While the free electron system has two equal Fermi surfaces with degenerate spin and the same Fermi momentum, the Rashba system has a chiral spin texture and two concentric Fermi surfaces at different Fermi momenta separated by $2k_{\text{R}}$.

When here and in the following the term ‘chirality’ is used, we refer to the alignment between spin and momentum, which for the eigenstates in the Rashba Hamiltonian are locked together as orthogonal to each other. The two resulting types of spin polarisation for a given carrier momentum have different eigenenergies and are referred to as the two opposite



Reprinted with permission from Ref. [143], P.C. Verpoort et al., J. Phys.: Condens. Matter 32, 355704 (2020). *Licensed under Creative Commons Attribution 4.0.*

Fig. 4.1 The energy dispersion of a free system with an additional Rashba interaction. Top: energy dispersion. Bottom: chiral spin structure at the Fermi surface with arrows indicating spin directions.

chirality species. We adopt the term ‘chirality’ from now on to refer to the two carrier types, although we note that some other terms exist in the literature, such as the ‘Rashba carrier type’ or the ‘helicity’ of carriers, which all refer to the same property.

A common experimental technique to ascertain the existence of Rashba coupling in the band structure of a solid-state system is through spin- and angular-resolved photoemission spectroscopy (often abbreviated as SARPES or spin-ARPES) [44], where the binding energy and in-plane momentum can be obtained from the exit angle and kinetic energy (as is done in standard ARPES measurements), while the spin of the emitted electron is determined using a spin detector. While several techniques exist to measure the spin, the most common method employs a Mott detector, which is based on the spin-dependent diffraction off heavy nuclei, an effect first studied by Mott [94]. The main limitation of SARPES is that it can only probe surface or near-surface states, yet it has effectively become the standard method for the investigation of spin-polarised energy dispersions, as it is capable of directly measuring the electron spin, momentum, and energy at the same time.

Another common technique for establishing the existence of a chiral spin texture at the Fermi level based on transport measurements is the observation of a magnetoresistance curve featuring weak anti-localisation (WAL) behaviour, which is sometimes also referred to as a

WAL ‘cusp’. The standard weak localisation (WL) behaviour present in all regular metallic samples below a certain temperature threshold arises due to quantum coherence effects, which makes closed-loop paths interfere constructively and results in an increase of the overall resistivity [3]. In contrast, in systems with a chiral spin texture, the phase acquired in a closed-loop path results in destructive interference and hence a reduced resistivity. Adding an external magnetic field results a destruction of the WAL effect, which manifests in the WAL cusp described by the Hikami-Larkin-Nagaoka (HLN) equation [61]. Reporting a cusp in the magnetoresistance curve can thus in some contexts be regarded as proof for the existence of Rashba coupling. It should however be mentioned that Rashba coupling is not the only source of WAL effects, as other dispersions with a chiral spin texture, such as surface states of topologically non-trivial systems, can also result in a WAL effect. Consequently, whether the observation of WAL in the magnetotransport curve can be seen as sufficient evidence for Rashba coupling has to be determined depending on the specifics of the system under consideration.

Rashba materials have attracted the interest of condensed-matter and solid-state physicists for a number of reasons in recent years. Their strong intrinsic spin-orbit coupling and the ability to control its direction in ferroelectric materials by flipping the ferroelectric polarisation through external electric fields have generated much interest in the context of designing devices for memory and logics based on spin currents – a field also well known as spintronics [153, 163]. Moreover, exciting novel effects have been predicted and observed in systems with strong Rashba coupling, including the universal intrinsic spin-Hall effect [126], the WL to WAL transition [92], or spin-based logical circuits [85, 128]. Moreover, non-centrosymmetric superconductors have been predicted to host Majorana zero-modes [120, 121, 134], hence making superconducting Rashba materials a potential candidate for discovering Majorana fermions, which in turn are much sought after to build robust devices for fault-tolerant quantum computing applications.

Having completed this preliminary introduction to Rashba systems, we can now proceed with defining the required methods from the field of non-equilibrium solid-state physics that are used in the subsequent two chapters.

4.2 Non-equilibrium physics in solid-state systems

The following two chapters will study Rashba systems in the context of non-equilibrium physics. This section therefore briefly summarises some basic concepts and theoretical background on non-equilibrium phenomena. Further details can be found in any standard textbook on solid-state physics [5, 86].

As mentioned in Chap. 1, the concepts of equilibrium and non-equilibrium are among the oldest and most fundamental concepts in statistical mechanics and date back to the early studies of heat transfer between bodies from the 19th century [54]. The definition of equilibrium, which initially was based on the exchange of energy and entropy, has since evolved and is nowadays more commonly defined as the stationarity of microscopic distribution functions, which is the definition that we shall work with in the following.

It is customary to describe the time evolution of microscopic states of many-body systems in and out of equilibrium using a semi-classical distribution function $f(\mathbf{k},\mathbf{r},t)$, whose value corresponds to the likelihood of a carrier occupying the state with momentum $\mathbf{k}$ and position $\mathbf{r}$ at time t and whose dynamics are governed by the semi-classical Boltzmann equation of motion,

$$\frac{\partial f}{\partial t}+\frac{\partial f}{\partial \mathbf{r}}\dot{\mathbf{r}}+\frac{\partial f}{\partial \mathbf{k}}\dot{\mathbf{k}}=\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{4.3}$$

The particle velocity is then normally obtained from the group velocity of a wave packet as $\dot{\mathbf{r}}={}^{\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k})}/_{\hbar}$, and the change of momentum, $\dot{\mathbf{k}}$, is given by the sum of all electromagnetic forces acting onto a particle in state $\mathbf{k}$. In the context of this work however, we will not be working with spatial gradients (i.e. $f(\mathbf{k},\mathbf{r},t)=f(\mathbf{k},t)$) of the distribution function or any electromagnetic forces, and hence we can significantly simplify this equation to a form that we shall work with in the following two chapters:

$$\frac{\partial f}{\partial t}=\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{4.4}$$

Hence, the change of the distribution function is entirely determined by the collisions of the charge carriers.

We will make additional assumptions regarding the form of the distribution function and its behaviour later in Chap. 6. In particular, we assume the distribution function to take the form of a free Fermi-Dirac equilibrium distribution, which is given by

$$f_{\text{Fermi}}(\mathbf{k})=\frac{1}{\mathrm{e}^{\beta(\varepsilon_{\mathbf{k}}-\mu)}+1}, \tag{4.5}$$

where $\varepsilon_{\mathbf{k}}$ is the energy dispersion of the respective charge carrier type, and μ is the chemical potential. While our research deals with non-equilibrium charge carrier distributions, our work is restricted to non-equilibrium in the form of chirality imbalances, such that the distributions of each carrier type will continue to exhibit a Fermi-Dirac equilibrium distribution.

Finally, one important approach taken in both Chap. 5 and Chap. 6 is the relaxation-time approximation, which is briefly outlined and described here and applied to the relaxation of chirality imbalances later on. A conventional attempt to evaluate the transport properties of conductors is to simplify the collision integral on the right-hand side of Eq. (4.3) and to replace it with a term that corresponds to a linear relaxation towards the equilibrium distribution f_0 with a time constant τ such that

$$\frac{\partial f}{\partial t}+\frac{\partial f}{\partial \mathbf{r}}\dot{\mathbf{r}}+\frac{\partial f}{\partial \mathbf{k}}\dot{\mathbf{k}}=-\frac{f-f_0}{\tau}=-\frac{f_1}{\tau}, \tag{4.6}$$

where we have defined f_1 to be the difference between the full distribution function and the corresponding equilibrium distribution, i.e. $f_1 = f - f_0$. The above approximation then typically yields an exponential relaxation of the distribution function with time constant τ to the equilibrium distribution. This approach can be used to derive expressions for relevant transport properties, such as the thermo-electric conductivity tensor. We shall take a fairly similar approach, albeit, instead of approximating the collision integral, we apply the relaxation-time approximation to electronic properties such as the chirality density C. We note that C, which will be defined again in Chaps. 5 and 6, is given as $C = n^- - n^+$, i.e. the difference between carrier densities of the two Rashba bands, $n^{\pm}$. This yields the differential equation

$$\frac{\mathrm{d}C}{\mathrm{d}t}=-\frac{C-C_{\mathrm{eq.}}}{\tau}. \tag{4.7}$$

By evaluating the collision integral and by expanding it in first order of $C - C_{\mathrm{eq.}}$ and $\frac{\mathrm{d}C}{\mathrm{d}t}$, this will allow us to derive an expression for the relaxation-time constant τ. This follows an approach taken by Yafet [158] that determines the spin-relaxation time constant from the electron-phonon interaction in a similar way. This procedure is further explained when applied in the next two chapters.

With this introductory section on non-equilibrium phenomena, methodology, and terminology completed, we are ready to continue our investigation of chirality relaxation in strongly Rashba-coupled systems in the next two chapters.

Chapter 5

Long-lived non-equilibrium states in Rashba systems

In this chapter, we summarise experimental observations of slowly decaying non-equilibrium effects in systems that exhibit strong Rashba spin-orbit coupling, where these effects manifest in anomalies of transport properties. After an in-depth discussion of existing experimental evidence, we analyse various theoretical frameworks that could potentially serve as a source of these observations, yet we demonstrate that due to a number of inconsistencies they all fail to truly explain the discovered novel non-equilibrium effects. Therefore, we put forward an alternative framework based on slowly decaying chirality imbalances. Based on this assumption, we model the dynamics of the Fermi energies and the resulting resistances.

As stated in the Preface to this thesis, all experimental evidence presented in this chapter has been obtained and kindly provided by V. Narayan and J.R.A. Dann, whose support is gratefully appreciated.

5.1 Experimental observations

This section introduces the key experimental observations that lead us to declare a novel previously undetected non-equilibrium effect in Rashba systems. The main focus of this chapter is on the material GeTe, which serves as the prime example and is used to define the novel non-equilibrium effect. In addition, we present one other system that displays non-equilibrium transport anomalies similar to those in GeTe and whose behaviour we attribute to the same effect, namely the $LaAlO_3/SrTiO_3$ system studied in Ref. [35].

There exist a few other potential candidate materials showing a similar behaviour that could be interpreted as the same effect, such as a Bi_2Te_3/Sb_2Te_3 vertical TI heterostructure [8], which features a non-equilibrium WAL cusp in its magnetoresistance similar to that of GeTe, $\mathrm{SmB_6}$ (Samarium-Hexaborite), where slowly decaying non-equilibrium states were first reported by Wolgast et al. [154], and finally $\mathrm{Au}_x\mathrm{Ge}_{1-x}$, where a strong non-equilibrium response of the magnetoresistance upon magnetic field-sweep with a clear rate dependence has also been observed [34]. There however remain many unresolved questions in those systems with respect to their properties, such as the exact origin of the WAL or the nature of the states carrying electrical current, and other sources for these observed effects cannot be ruled out with certainty, which is why they are not considered in this chapter.

More generally we note that, if the effect observed in experiment does in fact arise due to the existence of a strong Rashba spin-orbit coupling, there are plenty of other materials to study as potential candidates for realisation of the observed effect. These are of course the strongly coupled bulk-Rashba materials, including BiTeI and BiTeBr, as well as all materials that exhibit strong Rashba coupling on the interface, such as a InGaAs/InAlAs semiconductor interface, yet neither have these been studied systematically nor have observations been reported on them. Whether the same effect will be observable in other Rashba materials will obviously depend on several material-specific parameters, including, of course, the Rashba coupling strength (measured through the Rashba momentum k_R or the Rashba energy E_R) but also the carrier concentration, the effective mass, the resulting Fermi energy, and many other. In summary, there is great scope for extending the presented research to other materials of the Rashba class, yet such a systematic study is beyond the scope of this work.

5.1.1 Long-lived non-equilibrium superconductivity in GeTe

Introduction to the material

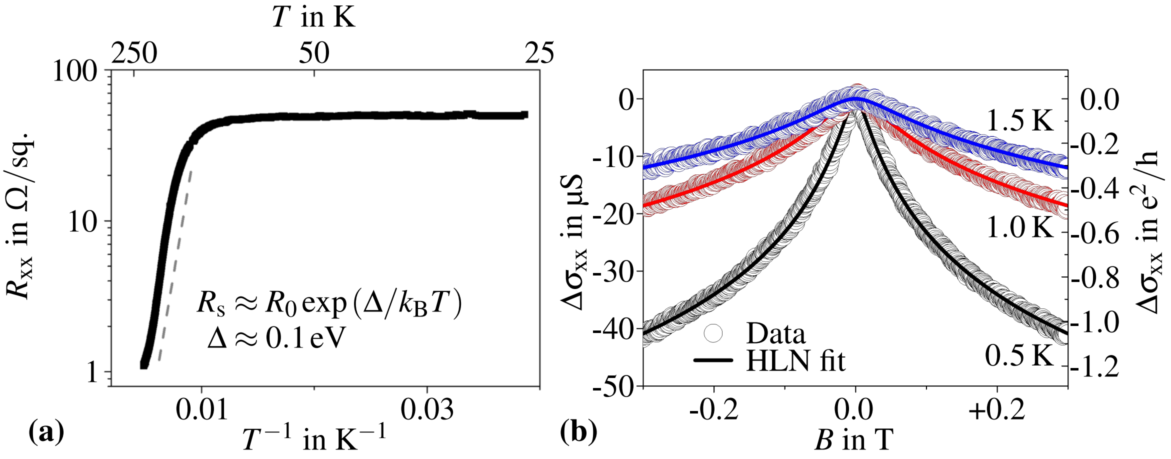
We commence by describing a novel non-equilibrium effect in GeTe, a narrow band-gap semiconductor that exhibits a giant bulk-Rashba spin-orbit coupling [108, 110, 112]. This bulk-Rashba effect arises from a non-centrosymmetric unit cell when the temperature is below its ferroelectric transition temperature of around[1] $T = 700\,\mathrm{K}$ so that the ions are displaced from their inversion-symmetric rocksalt configuration. While the nature of this ferroelectric transition is still under debate [28, 87, 127, 152, 157], this will not be of relevance to this study, as all reported experiments are well below the transition temperature, where the system can be expected to be fully in its rhombohedral phase with excitations to the inversion-symmetric cubic phase occurring at far higher temperatures. With a Rashba parameter of $\alpha_\mathrm{R} = 4.8\,\mathrm{eV\,\AA}$, a Rashba energy of $E_\mathrm{R} = 227\,\mathrm{meV}$, and a Rashba momentum of

[1] The exact value ranges between $T = 600\,\mathrm{K}$ and $T = 750\,\mathrm{K}$, and the precise value depends on the carrier concentration [127].

$k_\mathrm{R} = 0.09\,\text{Å}^{-1}$, GeTe ranks among the materials with the largest known Rashba momentum split [110] and will serve as our main candidate of investigation for non-equilibrium chirality relaxation. Due to the intrinsic Ge-vacancy doping of the material, holes are introduced into the valance band that push the Fermi energy just below the nodal crossing point of the Rashba band [110], giving the semiconductor its metallic properties [47, 48, 55, 73]. We note that, while GeTe is a hole-doped system, we always consider systems with positive effective mass in the following. Moreover, GeTe is known to become superconducting at temperatures of around $T = 100\,\mathrm{mK}$, where the exact value of T_c depends on the concentration of carriers [58], which in the high-quality molecular-beam-epitaxy (MBE)-grown samples presented here are dominantly provided by Ge vacancies.

Due to its strong Rashba coupling and its ferroelectric switchability, GeTe is currently investigated for its potential applications in spin-based logic schemes [85, 128] such as the Datta-Das spin transistor [38]. In addition to its potential applications in spintronics, GeTe is also a subject of investigation for its unconventional superconducting behaviour [58, 99]. Rashba systems are expected to exhibit Fulde-Ferrell-Larkin-Ovchinnikov [52, 76] (FFLO) phases, where the different Fermi momenta of the electron species give rise to Cooper pairs whose total momentum $\mathbf{q}$ is finite, which results in a spatially varying order parameter. As a consequence, while conventional superconductors normally show only a minor dependence on spatial variations of impurity concentrations, FFLO phases depend strongly on disorder [6, 64]. In addition to the possible FFLO physics present in GeTe, its ferroelectrically distorted unit cell also makes it a non-centrosymmetric superconductor, which have been theoretically shown to be of non-trivial topology if the p-wave pairing potential is larger than the s-wave potential, in which case they can host Majorana zero-modes in their edge states or in their vortex cores [120, 121, 134]. We note that all these effects are intriguing and make the GeTe material particularly complex to study. Our study of the novel non-equilibrium phenomena in GeTe however focusses on the normal-state behaviour, as this is less likely to be affected by FFLO physics and p-wave superconductivity.

The full band structure of GeTe has been investigated using density-functional theory (DFT) calculations [43, 110, 82, 115] and has been confirmed using SARPES [74, 82, 115]. This highlights a more complex band structure than described earlier in Fig. 4.1 due to a star-shaped Fermi surface in the lower Rashba band as well as regions of low effective mass and high density of states above the nodal point. While these effects would also be interesting to discuss in the context of the present work, we shall, for reasons of simplicity, restrict ourselves to approximating the GeTe band structure as a standard rotationally symmetric free energy dispersion with an additional Rashba term as introduced in Sec. 4.1 and visualised in Fig. 4.1. This simplification may of course affect the response of the system to the non-



Reprinted with permission from Ref. [100], V. Narayan et al., Phys. Rev. B 100, 024504 (2019). © *2019 American Physical Society.*

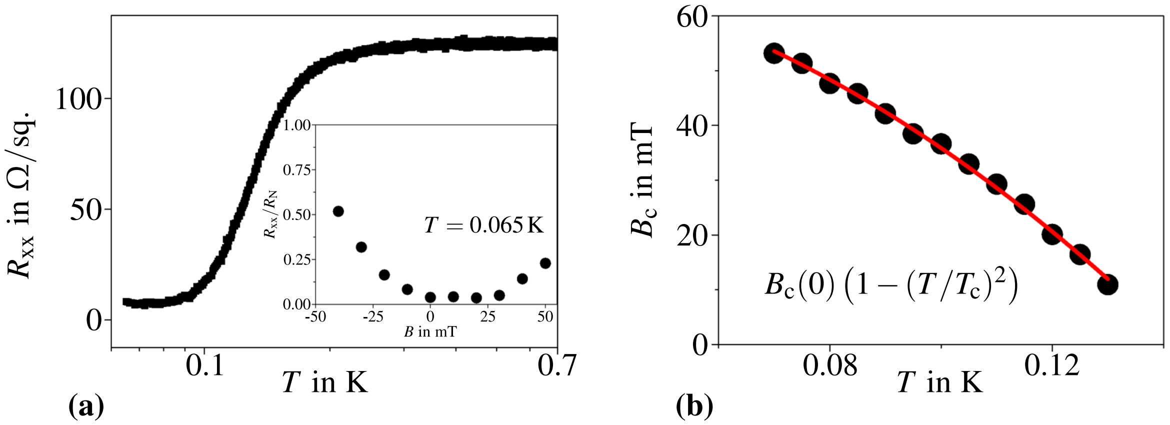
Fig. 5.1 **(a)** The GeTe films show activated and therefore semiconducting transport for $T > 100$ K, where a fit of the exponential growth yields a band gap parameter of $\Delta = 0.1$ eV. For $T < 100$ K, the resistivity saturates and becomes effectively constant, suggesting that conductance is dominated by metallic 2D surface states. **(b)** Measurements of the longitudinal conductivity with varying magnetic field strength feature a WAL cusp at zero field. The cusp can be described well by the HLN formula valid for 2D systems [61]. $\sigma_{xx} \equiv (L/W)\, R_{xx}/(R_{xx}^2 + R_{xy}^2)$, where R_{xx} and R_{xy} are the longitudinal and Hall components of resistance, respectively. Data obtained from Sample 1.

equilibrium carrier distributions that we consider in our model in the following sections, yet we do not expect the estimation of relaxation times and the fundamental underlying nature of the non-equilibrium effect to be much dependent on the precise band structure.

Characterisation of the sample

GeTe films of 18 nm thickness are grown onto Si(111) substrates using molecular-beam epitaxy (MBE). These are then fabricated into Hall bars to measure the longitudinal conductance (note that the Hall conductance was also measured but this is not discussed here, for further details check Ref. [100]). This section reports results from two GeTe samples patterned from the same wafer, in the following referred to as Samples 1 and 2. The equilibrium properties, as determined by the behaviour of the sample in the infinite-time limit for constant magnetic field, are characterised by transport measurements reported in Figs. 5.1, and 5.2.

The temperature dependence of the resistivity shown in Fig. 5.1(a) displays activated behaviour down to $T = 100$ K, where a fit of the exponential growth of the resistivity suggests a band gap of $\Delta = 100$ meV, which is in good agreement with ARPES measurements [74] of GeTe of $\Delta = 60$ meV. Below a temperature of $T = 100$ K, the resistivity saturates at a



Reprinted with permission from Ref. [100], V. Narayan et al., Phys. Rev. B 100, 024504 (2019). © *2019 American Physical Society.*

Fig. 5.2 **(a)** The sample undergoes a broad superconducting transition between $T = 0.1$ K and 0.2 K below. We note that R_{xx} does not vanish fully, which may possibly arise due to a positive voltage offset during the measurement (for further details check Ref. [100]). Inset: superconductivity is suppressed when the sample is cooled in the presence of a constant magnetic field perpendicular to the plane of the film. Note that this data suggests a possible field offset of $B = 15$ mT, which can occur due to trapped magnetic flux in the external superconducting magnet. **(b)** Plotting the critical field B_c as a function of T allows to obtain $B_c(0\,\mathrm{K}) = 70$ mT and $T_c = 140$ mK from extrapolation. For this extrapolation the superconducting transition is defined for $R_{xx} = R_N/2$, where R_N is the normal-state resistance. Data obtained from Sample 2.

constant level, which suggests that metallic 2D surface states exist that dominate transport below that temperature.

This conjecture of 2D transport is further corroborated by the observed WAL cusp in the resistivity at $T < 1$ K, as shown in Fig. 5.1(a). The strong Rashba spin-orbit coupling gives rise to quantum corrections of transport properties, as outlined in Sec. 4.1, which results in an enhancement of the electrical conductivity for non-zero external magnetic fields, i.e. $\Delta\sigma_{xx} = \sigma_{xx}(B) - \sigma_{xx}(0)$. A functional fit of the HLN formula valid for 2D systems [61] is in good agreement with the obtained data and therefore suggests that transport at low temperatures is of 2D type. In fact, the existence of 2D metallic modes is consistent with spectroscopic measurements in GeTe [82]. We note that, due to the apparent 2D nature of the conductance, the electric resistivity is in all plots reported as 'sheet resistance' given by $R \times W / L$, where R is the resistance obtained from a four-terminal setup, and W and L are the width and length of the Hall bar. Finally, the onset of superconductivity at $T = 0.2$ K and its suppression through a perpendicular magnetic field are shown in Fig. 5.2.

In summary, we have so far presented a strongly Rashba-coupled semiconductor with metallic 2D conducting states that are the sole carrier of transport below $T = 100$ K. This

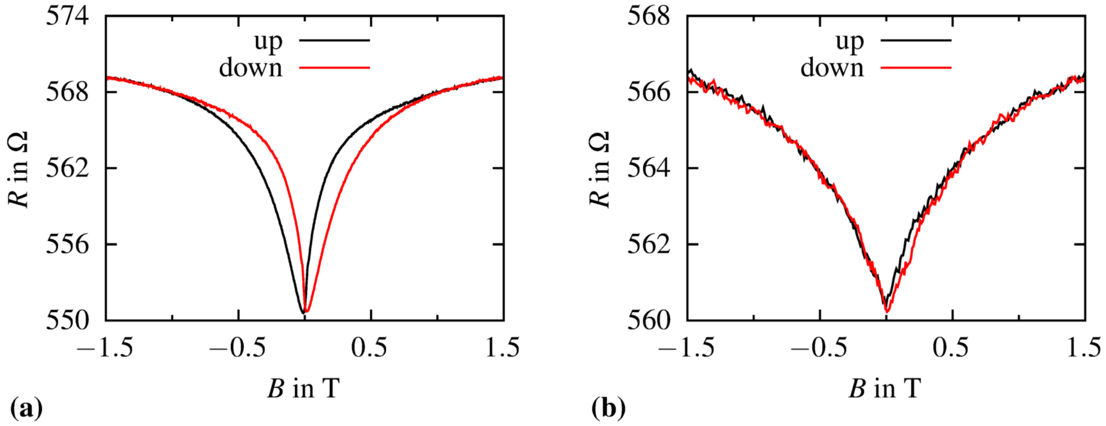


Fig. 5.3 Non-equilibrium magnetoresponse of GeTe for a continuously swept external magnetic field. **(a)** Resistivity when measured under alternating current (AC). The traces obtained from up and down sweeps follow different curves. This measurement is performed using frequency $f = 77\,\mathrm{Hz}$, current $I = 100\,\mathrm{nA}$ and sweep rate $\dot{B} = 10\,\mathrm{T\,h^{-1}}$. **(b)** Resistivity when measured under direct current (DC). The non-equilibrium effect observed for AC ceases when switching to DC measurement. This measurement is performed using current $I = 2\,\mu\mathrm{A}$, and sweep rate $\dot{B} = 10\,\mathrm{T\,h^{-1}}$.

system features a WAL cusp, indicative of its strong Rashba coupling as well as superconducting transport properties around $T = 100\,\mathrm{mT}$, which is suppressed by external magnetic fields larger than the (zero-temperature) critical field of $B_\mathrm{c}(0\,\mathrm{K}) = 70\,\mathrm{mT}$.

Slowly decaying non-equilibrium state

We now continue by examining the non-equilibrium magnetoresistance curves of this material. In the following, we report an extremely long-lived non-equilibrium state of the system, which is observable due to its salient transport properties, and which relaxes on macroscopic timescales of several minutes. In this section, we only described the observed properties, and we postpone the consideration of existing frameworks to explain these results and the presentation of our novel interpretation to Secs. 5.2 and 5.3.

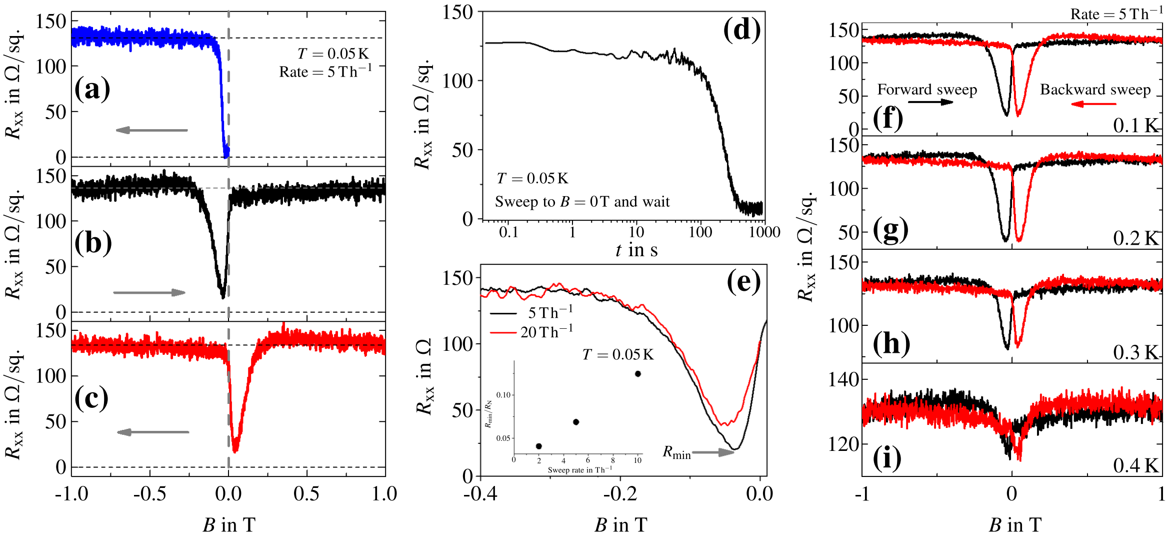
We set off by reporting the normal-state (i.e. non-superconducting) non-equilibrium behaviour of the sample in Fig. 5.3. When measured with alternating current (AC), the magnetoresistance (R as a function of B) follows two different curves depending on the sweep direction of the magnetic field. The difference in resistivity between the continuously swept curve and the equilibrium curve decays exponentially with time over a period of several minutes. Strikingly, the non-equilibrium effect does not occur when the same measurement with the same sample is performed with direct current (DC). This observation will serve later as one important argument to support the proposition that this effect cannot be explained by existing frameworks, such as nuclear spin effects or magnetocaloric effects (see Sec. 5.2).

We note that evidence of this non-equilibrium effect in the normal state of GeTe was already reported previously by Narayan et al. in Ref. [99], where the novelty of this effect was not yet fully understood. In this publication it is presented as a hysteretic effect, although the Supplementary Information of this article presents a graph showing the dependence of the effect on the sweep rate. We recall that hysteresis is defined as a ‘rate independent memory effect’ [144] and is an equilibrium phenomenon that does not decay over time. As the effect observed in GeTe decays with time and, therefore, is not an equilibrium effect (as defined in Sec. 4.2), it does not fall into the common definition of hysteresis. Furthermore, we stress that the magnetoresistance traces reported in Fig. 5.3 should not be confused with the so-called ‘butterfly hysteresis’ observed in magnetic Dirac materials [20, 98, 136, 154] because, again, the observed effect is not hysteretic but of non-equilibrium nature. As already discussed in Ref. [99], the curves in Fig. 5.3 are asymmetric and can therefore not be explained by a mere field shift (although this could in theory be created by an remnant hysteretic magnetic field following a hysteresis loop, yet such fields are also excluded, as further discussed in Sec. 5.2).

Next, we report evidence of the non-equilibrium effect that involves superconducting transport behaviour, where the presented data is summarised in Fig. 5.4. While the onset of a magnetic field quickly destroys the superconductivity, as seen in Fig. 5.4a (and outlined before in Fig. 5.2), a continuously swept field can create a drop to almost zero resistivity at magnetic field strengths higher than the equilibrium upper critical field, $B > B_\mathrm{c}$, as is visible in Figs. 5.4b and 5.4c. Moreover does this non-equilibrium resistance drop vanish at zero field, where the equilibrium superconductivity would be expected to occur. We attribute this to the existence of a second non-equilibrium superconducting state. This state decays slowly over the period of several minutes, as is evident from Fig. 5.4d, and even occurs above the critical temperature T_c of the equilibrium superconductor, as observed in Fig. 5.4(f)–(i). The absence of a drop to absolute zero resistance could be interpreted as a competition between the timescales for the superconducting transition and the decay of the non-equilibrium state.

Summary

In summary, we have presented a novel non-equilibrium effect in GeTe, which manifests in transport properties for both the normal and the superconducting behaviour. This non-equilibrium state decays over the course of several minutes, hence is extremely long lived. Its existence seems to be strongly dependent on the rate at which the external magnetic field is ramped. Moreover, its resistance drop occurs even above the upper critical field, B_c,



Reprinted with permission from Ref. [100], V. Narayan et al., Phys. Rev. B 100, 024504 (2019). © *2019 American Physical Society.*

Fig. 5.4 Evidence of non-equilibrium effect with superconducting transport properties. **(a)** Starting in equilibrium in the superconducting state, an external magnetic field destroys the superconductivity at $B = -15$ mT, and the system transitions to its normal state. **(b)** However, when starting at $B = -1$ T and slowly ramping the field up to $B = 1$ T, the resistance drops almost all the way to zero. This starts to occur at $B = -250$ mT, i.e. a field much higher in strength than the apparent critical field, B_c, observed in (a) and Fig. 5.2. Moreover, this drop in resistance vanishes sharply at zero field, leaving the system at $R_{xx}(B = 0) \approx R_N$. **(c)** The mirror image of (b). Grey arrows indicate sweep direction of the magnetic field. **(d)** Stopping the sweep at zero field allows to observe the decay of the non-equilibrium state. It persists for ≈ 100 s before relaxing to equilibrium over 300 s. **(e)** The height of the resistance drop is reduced by increasing the sweep rate. This indicates that an optimal value exists for the sweep rate, $\mathrm{d}B/\mathrm{d}t$, below 5 T h^{-1}. **(f)** – **(i)** While the equilibrium superconductivity vanishes entirely for $T > 200$ mK (compare Fig. 5.3), the non-equilibrium resistance drop persists up to $T = 400$ mK. Data obtained from Sample 1, except for inset in (e), which is obtained from Sample 2.

and critical temperature, T_c, of the equilibrium superconducting state, hence is not directly contingent on the equilibrium superconductivity of the system.

5.1.2 Transient superconductivity at LaAlO$_3$/SrTiO$_3$ interface

Having completed a report of the salient non-equilibrium properties of GeTe, the main material studied in this work, we continue by comparing this with the behaviour of one other material that shows comparable non-equilibrium properties.

The $LaAlO_3/SrTiO_3$ system

The material presented in this section is a two-layered $LaAlO_3/SrTiO_3$ system that features non-equilibrium transport properties under a continuously swept magnetic field involving both normal-state and superconducting behaviour, of which experimental evidence was recently reported by Daptary et al. [35]. In summary, the material features (1) a 2D electronic system on the $LaAlO_3/SrTiO_3$ interface with Rashba spin-orbit coupling (evidence of which is obtained through WAL effects in transport measurements), (2) a slowly decaying non-equilibrium magnetoresistance curve in the normal state, and (3) a superconducting state that is accessible through the application of a continuously swept external magnetic field.

While the article offers an explanation for the onset of a decaying superconducting state under continuous field sweep based on a reduction of the magnetic field strength of a neighbouring magneticly ordered system, we believe that this is insufficient to explain the observation, and we put forward our model as an alternative attempt to explain the effect.

Previous interpretation

As outlined by Ref. [35], it is believed that the d_{xy} levels of the d-t_{2g} orbitals of the Ti atoms at the interface are saturated first due to the broken crystallographic inversion symmetry at the interface, which give rise to ferromagnetism [106] because of strong on-site Coulomb repulsion. Once the electron density reaches a certain level, the d_{xz} and d_{yz} levels are occupied, which have a higher mobility and constitute the main carrier of transport in the system. Ref. [35] then claims that the observed non-equilibrium superconductivity can be explained based on a derived model that, under the assumption of a long spin-lattice and spin-spin relaxation timescale of 100 s–200 s, the magnetic field of the ferromagnetic state in the d_{xy} orbitals is sufficiently destroyed to no longer suppress the superconducting state in the d_{xz} and d_{yz} orbitals.

While this explanation is interesting, several inconsistencies can be identified. First, it falls short of explaining the normal-state non-equilibrium behaviour, which the article refers to as hysteresis, despite the identification of its slow decay with time: “The hysteresis is time dependent and relaxes exponentially to an equilibrium value over a time scale of a few hundreds of seconds.” [35] We stress again that this identification of a non-equilibrium effect as a hysteretic effect is in disagreement with the conventional definition of hysteresis as an equilibrium memory effect that does not change with time [144]. As an alternative explanation for this observation, magnetocaloric effects are considered: “Although hysteresis in magnetoresistance in the low doping regime has been seen previously in $LaAlO_3/SrTiO_3$ heterostructure devices and was taken to indicate the presence of ferromagnetic domains in the system..., there is a growing concern in the community that it might also have contributions from induction effects due to fast magnetic field sweeps." [35] This argument is inconsistent, as heating effects arising from induced current should not be dependent on the direction of the field sweep. Moreover, it is unlikely to be the source of the slow decay of the magnetoresistance anomaly since the thermalisation timescale of the sample can be expected to be much less than the observed relaxation timescales, as is also the case for the GeTe samples from the previous section. Second, this explanation is based on the assumption that the localised ferromagnetic moments in the d_{xy} orbitals follow Bloch-spin dynamics with spin-lattice and spin-spin relaxation timescales of 100 s–200 s. It is questionable whether such long relaxation timescales can generally be expected from these spins. For comparison, a study of conduction-electron spin resonance in zinc-blende GaN thin films [50] has revealed a spin-lattice relaxation time on the order of 6×10^{-5} s at $T = 10\,\mathrm{K}$ and $B = 1.8\,\mathrm{mT}$, where the resonance is attributed to non-localised electrons in a band of shallow donors arising from N vacancies. Arguably, the d_{xy} levels on the LaAlO$_3$/SrTiO$_3$ interface are more localised, yet it is unclear whether this suffices to justify such vastly different relaxation times.

Summary

The inconsistencies of the model presented by Ref. [35] opens up the space for alternative interpretations, including a model put forward further below in Sec. 5.3. While we note that we are unable to understand the specifics of the LaAlO$_3$/SrTiO$_3$ system to the extent that we can in GeTe, as parameters such as the Rashba momentum k_R are unknown, we believe that this material could be considered as another example of the same novel effect reported for GeTe in the previous section.

5.2 Other possible explanations

A multitude of mechanisms could be suggested as a potential source for the non-equilibrium effects in the mangetoresistance of the two systems outlined above, which could explain their unconventional magnetotransport properties and the slowness of the relaxation back to equilibrium. In this section, we examine the most relevant ones, in particular ferromagnetism, magnetocaloric effects, trapped flux, and nuclear spins. We are able to demonstrate why these cannot serve as a way to explain the observed behaviour.

Ferromagnetic hysteresis: First and foremost, we reiterate that the observed effects cannot arise due to hysteresis stemming from ferromagnetic alignment of magnetic moments in the sample. It is easy to misinterpret the magnetoresistance curves as a hysteresis effect, and in fact several works erroneously report the observation of hysteresis. The fact that this is incorrect can be understood by acknowledging – again – that hysteresis is an equilibrium effect that does not change with time, whereas the effects discussed in the previous sections clearly decay with time and are thus of non-equilibrium nature.

Magneto-caloric effects: The perhaps most prominent alternative explanation attributes the non-equilibrium response of those systems under a ramped external magnetic fields to magnetocaloric effects, i.e. the exchange of heat of the system with the varying magnetic field, which would result in a change of temperature. This could either arise from adiabatic cooling of some intrinsic spin degree of freedom in the system as spin ordering is lost when the external field ceases. Equally, the change of the magnetic field strength, which induces Eddy currents, could lead to magnetic heating. This can however be excluded, at least for the GeTe samples, where the thermalisation timescale has experimentally been verified to be significantly shorter than the observed non-equilibrium timescale in the experiments. The exchange of heat between the sample and its bath results in an abrupt and instantaneous return to the base temperature of the cryostat, whereas the non-equilibrium magnetotransport properties persist over a timescale of several minutes, even when the magnetic field sweep is entirely stopped.

Trapped flux in the superconducting magnet: The external magnetic field that is used to obtain the magnetoresistance curves is supplied via a superconducting magnet. It is possible that magnetic flux could be trapped in the superconducting coil, resulting in an offset of the magnetic field strength, which can in some cases decay on very long timescales. This however would only be an offset in B and cannot explain the dynamical drop in resistivity above T_{c}. Additionally, the magnetic field offset from such trapped fluxes would correspond to a paramagnetic correction to the external field. This is clearly inconsistent with the experimental observations, which would suggest a diamagnetic correction since the minimum of R_{xx} in Fig. 5.3 appears before crossing $B = 0\,\mathrm{T}$.

Nuclear spin dynamics: Another potential candidate for the source of the observed effects are dynamics arising from nuclear spins, in particular because the relaxation timescales associated with nuclear spin dynamics are typically quite long and can in principle be on the order of the ones observed here in experiments. However, two shortcomings cast doubt on nuclear spins as a source: 1) the alignment of nuclear spins only results in the onset of an additional weak magnetic field, also known as the Overhauser field, and it is unclear how this field, which is typically rather weak and usually on the order of a few mT [135], could serve as a source for the non-equilibrium magnetoresponse, especially in those examples where $T > T_{\mathrm{c}}$; 2) it is unclear how the change of the external magnetic field, which apparently is the driver of the non-equilibrium response of the samples, could affect the polarisation of the nuclear spins. Nuclear spins are affected by the the hyperfine interaction with the conduction electrons. This hyperfine interaction can only, if at all, depend on B, and not on its temporal change, ${}^{\mathrm{d}B}\!/\!{}_{\mathrm{d}t}$.

Finally, it is worth noting that all the above-mentioned effects, which could act as candidates for the source of the observed non-equilibrium magnetotransport properties, fail to offer an explanation for the discrepancy between the AC and DC responses in GeTe, which are presented in Fig. 5.3. This different behaviour under AC or DC operation indicates that the observed effects could in fact be manifested in a non-equilibrium charge carrier distribution, which cannot arise due to any of the effects listed in this section. This idea that the non-equilibrium response of the system is rooted in a non-equilibrium distribution of the carriers will be taken up and discussed in more detail in the next section.

5.3 Explanation of the observations based on slow chirality relaxation

Having presented experimental evidence of slowly decaying non-equilibrium magnetotransport properties in systems with Rashba coupling and having ruled out all conventional ways of explaining these findings, we will now develop a model that aims to explain the observations based on slow electronic chirality relaxation based on the Rashba energy dispersion.

5.3.1 Introduction

We start by outlining the basic principles that guide our development of a new model theory. The following main observations will need to be incorporated when finding a new model: (1) the system is subject to strong Rashba coupling, (2) the system can host non-equilibrium charge carrier configurations that decay slowly and are created through a varying external magnetic field, and (3) these non-equilibrium charge carrier configurations lead to different transport properties.

Points (1) and (3) are evident from the experimental evidence presented in Sec. 5.1, whereas point (2) might require further clarification. This additional assumption that the non-equilibrium nature of the system arises from charge carrier redistributions is based on the observation that the dynamical magnetoresistance curves only occur when the system is measured under AC but vanish when the system is measured under DC, as shown in Fig. 5.3. This observed dependency on the current frequency allows us to conclude that the non-equilibrium state of the system arises due to a non-equilibrium distribution function of the charge carriers. This conclusion is based on the interpretation that driving a nonalternating current through the system could cause carriers entering the sample from the leads to populate the electron distribution function in a way that relaxes the non-equilibrium state of the system, whereas an alternating current of sufficiently large frequency will not move the carriers enough to allow this to happen.

Based on the assumptions (1)–(3), we develop a theoretical framework in the next section that aims to model the experimental observations reported in Sec. 5.1.

5.3.2 Modelling of chirality imbalance relaxation

The model presented in this section is based on the Rashba energy dispersion introduced in Sec. 4.1 but with an additional coupling to an external magnetic field **B**. This is described by the Hamiltonian

$$H = \frac{\hbar^2 \mathbf{k}^2}{2m} - \alpha_\mathrm{R} \boldsymbol{\sigma} \cdot (\mathbf{r}_\mathrm{SO} \times \mathbf{k}) + g\mu_\mathrm{B} \boldsymbol{\sigma} \cdot \mathbf{B}\,, \tag{5.1}$$

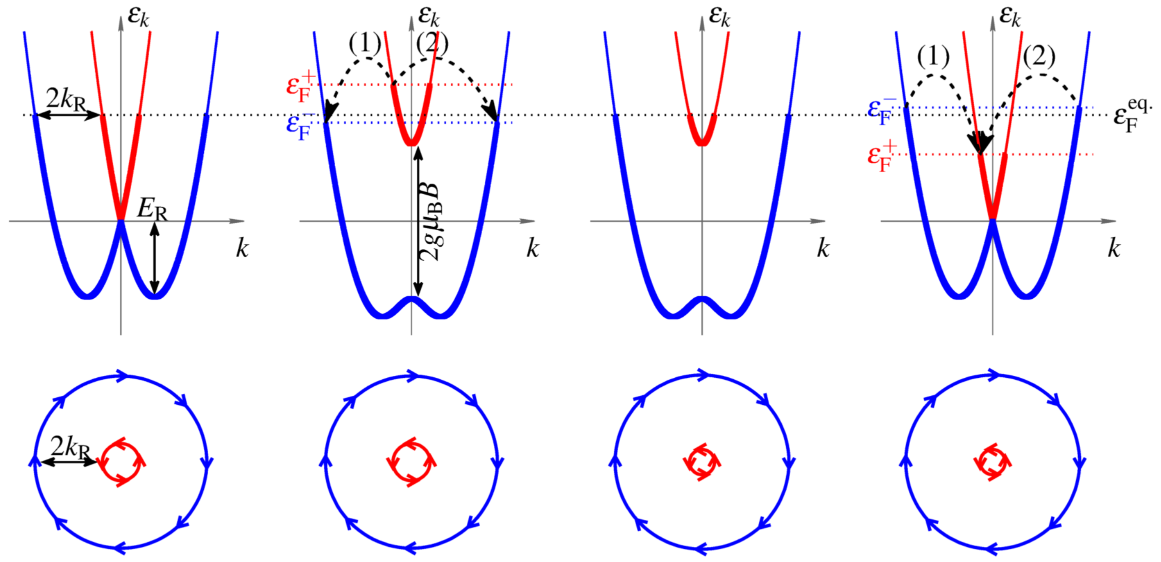
where the second term on the right is the Zeeman energy due to the external magnetic field $\mathbf{B} = B\hat{z}$, which is taken to point along $\mathbf{r}_\mathrm{SO}$, g is the Landé g-factor, and μ_B is the Bohr magneton. We consider the case $\mathbf{B} \parallel \mathbf{r}_\mathrm{SO}$ for concreteness but note that configurations where **B** is not perfectly parallel to $\mathbf{r}_\mathrm{SO}$ do not affect the results qualitatively. Eq. (5.1) then gives the energy dispersion

$$\varepsilon_k^{\pm} = {\hbar^2 \mathbf{k}^2}/{2m} \pm \sqrt{(g\mu_\mathrm{B} B)^2 + (\alpha_\mathrm{R} \mathbf{r}_\mathrm{SO} \times \mathbf{k})^2}\,, \tag{5.2}$$

where the $+$ $(-)$ superscript refers to the inner (outer) Rashba band, respectively. The dispersions with and without B field and the resulting k-dependent spin alignments on the Fermi surface are shown in Fig. 5.5(a) and Fig. 5.5(c), respectively. We note that the Rashba dispersions in those figures are similar to the one shown in Fig. 4.1, however with a gap introduced at zero field in Fig. 5.5(c).

In a typical electronic system, the relaxation of a non-equilibrium carrier population back to equilibrium can be expected to happen on very short timescales, typically on the order of a picosecond. In a Rashba system, relaxation requires scattering events that induce transitions between the Rashba bands. As indicated in Fig. 5.5(b) and Fig. 5.5(d), these can broadly be split up into two types: (1) small-angle scattering events that change the momentum only marginally but have to flip the spin and (2) large-angle scattering that reverses the momentum but leaves the spin unchanged. Clearly, the entirety of all scattering events includes cases in between (1) and (2), and these will be discussed further in Chap. 6.

The argument of the long timescales for this relaxation process is established by considering the energy-momentum conservation for phonons and the nature of small-angle


ε_k
$2k_\mathrm{R}$
E_R
k
(1)
(2)
ε_F^+
ε_F^-
$2g\mu_\mathrm{B}B$
$\varepsilon_\mathrm{F}^\mathrm{eq.}$

$B = 0$ **(a)** equilibrium $B \neq 0$ **(b)** non-equilibrium $B \neq 0$ **(c)** equilibrium $B = 0$ **(d)** non-equilibrium

Fig. 5.5 Charge carrier configurations in Rashba systems when sweeping a magnetic field. The four snapshots correspond to the instantaneous non-equilibrium and the relaxed equilibrium carrier configurations after the B field was turned on and off. Top: energy dispersion with bold lines representing occupied states. Bottom: chiral spin structure at Fermi surface with arrows indicating spin directions. **(a)** The spin-orbit coupling and lack of inversion symmetry lift spin degeneracy, resulting in two parabolic dispersions centred symmetrically away from $k = 0$. Consequently, Rashba systems have two concentric Fermi surfaces with opposite spin helicity that are separated in momentum space by $2k_\mathrm{R}$. **(b)** Under an applied B field a Zeeman gap $\varepsilon_\mathrm{Z} = 2g\mu_\mathrm{B}B$ opens at momentum $k = 0$, pushing the inner band up in energy and lowering the outer band. The panel shows a non-equilibrium situation in which the bands have unequal Fermi levels. Equilibrium is restored via the inter-band transitions. The transitions must either involve a spin-flip or a reversal of momentum (labelled 1 and 2, respectively). **(c)** When equilibrium is achieved the inner Fermi surface has reduced in size and the outer one has grown as compared to the $B = 0\,\mathrm{T}$ case. **(d)** When the field is swept back towards $0\,\mathrm{T}$, the configuration of the bands is opposite to that in (b), with the outer band at a higher Fermi level. Equilibration is achieved via processes such as 1' (spin reversal) and 2' (momentum reversal).

scattering. In a system with strong Rashba coupling, the value of k_R can be much larger than the thermal phonon momentum scale at low T. This means that inter-band transitions induced by the electron-phonon interaction are prohibited for simple reasons of energy-momentum conservation. Furthermore, scattering with charged impurities and inter-carrier Coulomb scattering is dominated by the transfer of small momenta, which for inter-band transitions is incompatible with the Rashba coupling for two reasons: 1) the split of the Fermi surfaces in momentum space by $2k_\mathrm{R}$ acts as a small-momentum cutoff and 2) the spinor overlap vanishes for transitions with the smallest possible momentum transfer. This is discussed in greater detail in Chap. 6.

We therefore conjecture that Rashba systems display slowly decaying non-equilibrium carrier populations that are similar to the ones shown in Fig. 5.5(b) when T is lower than a threshold below which the phonon bath is unable to provide the momentum required to induce inter-band transitions.

We will investigate the time evolution of this system under the influence of a time-dependent magnetic field as depicted in the transition from Fig. 5.5(a) to Fig. 5.5(d). If the field is swept sufficiently slowly, the adiabatic theorem holds, which implies that the time-dependent eigenstates are approximately equal to the instantaneous eigenstates (sometimes referred to as 'snap-shot eigenstates') of the Hamiltonian and that transitions into other states only occur through terms of a higher-order perturbation expansion in the slowness of the field sweep [114]. Consequently, as the energy of the instantaneous eigenstates in the inner (outer) Rashba band is increased (decreased), the respective Fermi energies of the two bands follow this motion, which results in a detuning of the Fermi energies of the two bands. At the same time, the Fermi momenta of the two bands remain unchanged. By applying a time-dependent B field, the system is therefore brought into a state, where the Fermi energies $\varepsilon_\mathrm{F}^\pm$ of the two Rashba bands are not equal to each other, as is shown in Fig. 5.5(b) and Fig. 5.5(d).

We now attempt to model the Fermi energy dynamics described above and incorporate the effects of a time-dependent magnetic field as well as a slow exponential decay of the chirality to its equilibrium value. The Fermi energies of the two Rashba bands, $\varepsilon_\mathrm{F}^\pm$, are a function of B as well as the occupation number of the respective band, $n^\pm$, and hence their time evolution is governed by the differential equation

$$\frac{\mathrm{d}\varepsilon_\mathrm{F}^\pm}{\mathrm{d}t} = \frac{\partial \varepsilon_\mathrm{F}^\pm}{\partial B}\frac{\partial B}{\partial t} + \frac{\partial \varepsilon_\mathrm{F}^\pm}{\partial n^\pm}\frac{\partial n^\pm}{\partial t}. \tag{5.3}$$

The first term drives the Fermi energies out of equilibrium, whereas the second term allows the charge carriers to relax back to two equal Fermi energies. We note that in the following steps we will always assume that the equilibrium Fermi energy of the system is above the nodal point, i.e. $\varepsilon_\mathrm{F}^\mathrm{eq.} > 0$.

In order to model $\frac{\partial n^\pm}{\partial t}$, we use a relaxation-time approximation with time constant τ. We define the total carrier density, $n = n^+ + n^-$, and the chirality density, $C = n^- - n^+$. The relaxation-time approximation then takes the form

$$\frac{\partial C}{\partial t} = -\frac{C - C^\mathrm{eq.}}{\tau}, \tag{5.4}$$

as explained in Sec. 4.2. Furthermore, we can assume that the magnetic field changes at a constant rate, hence

$$\frac{\partial B}{\partial t} =: \dot{B} = \text{const.} \tag{5.5}$$

This allows us to derive an ordinary coupled differential equation for $\varepsilon_{\mathrm{F}}^{+}$ and $\varepsilon_{\mathrm{F}}^{-}$,

$$\begin{aligned}\frac{\mathrm{d}\varepsilon_{\mathrm{F}}^{\pm}}{\mathrm{d}t} =& \frac{b\dot{b}}{-2E_{\mathrm{R}} \pm \sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}} \\ &\pm \frac{1}{2\tau}\left(1+\frac{2E_{\mathrm{R}}}{-2E_{\mathrm{R}} \pm \sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}}\right) \times \Bigg(\varepsilon_{\mathrm{F}}^{-}-\varepsilon_{\mathrm{F}}^{+}+2\sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}} \\ &\qquad -\sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{+}}-\sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{-}}\Bigg)\,,\end{aligned} \tag{5.6}$$

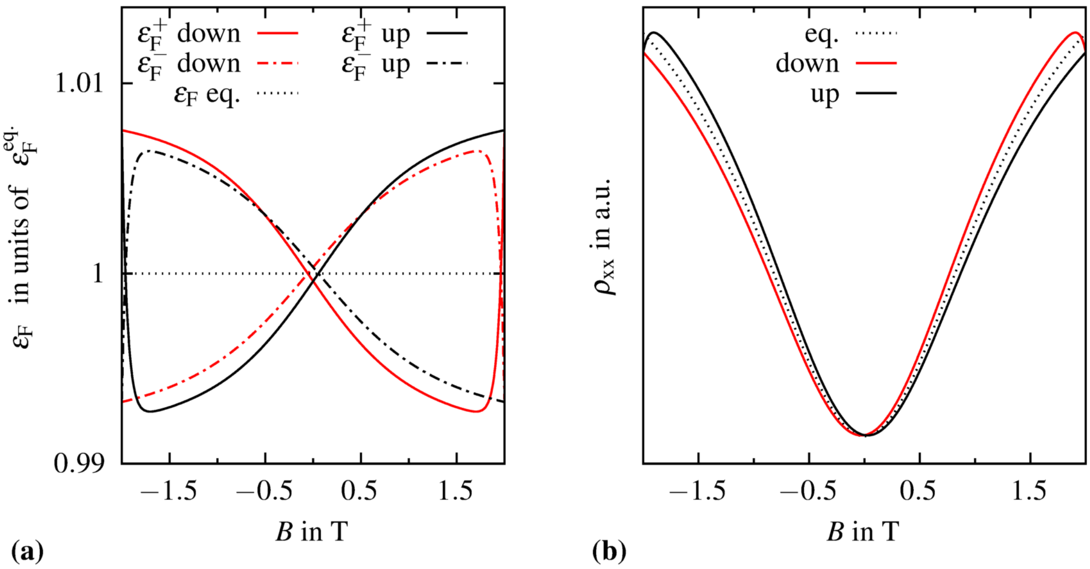
where $b := g\mu_{\mathrm{B}}B$. The steps to obtain this equation from Eqs. (5.3), (5.4), and (5.5) are outlined in App. B.1. We can integrate Eq. (5.6) and obtain the dynamical Fermi energies, $\varepsilon_{\mathrm{F}}^{\pm}(t)$, which are plotted in Fig. 5.6 for $\tau = 40\,\mathrm{s}$ in accordance with the experimental observations[2], $\dot{B} = 1\,\mathrm{T\,h^{-1}}$, $\varepsilon_{\mathrm{F}} = 0.075\,\mathrm{meV}$, and $E_{\mathrm{R}}/\varepsilon_{\mathrm{F}} = 0.01$, where the latter two parameters have been adjusted such that the resulting resistance graph qualitatively matches the results from Sec. 5.1.

In addition to the report of the Fermi energies, one can also calculate the resistivity. In theory this would require a careful analysis of the WAL quantum corrections to transport and how these would be affected by a Fermi level imbalance. This is however much too complex for the simple model we are employing here, and so we work with classical transport theory, for which the difference in conductivity per carrier of the two Rashba bands arises purely due to the different effective masses at the Fermi level. In Fig. 5.6(b) we report $\rho_{\mathrm{xx}}(B) \sim 1/(\sigma_{\mathrm{xx}}^{+}+\sigma_{\mathrm{xx}}^{-})$ as derived from the conductivities $\sigma_{\mathrm{xx}}^{\pm} \propto \int \mathrm{d}^3k\,(\partial\varepsilon_{\mathbf{k}}^{\pm}/\partial k)^2\,\delta(\varepsilon_{\mathbf{k}}^{\pm}-\varepsilon_{\mathrm{F}}^{\pm})$.

We stress that the results reported in Fig. 5.6 are very qualitative and should not be quantitatively compared to the experimental observations in Fig. 5.3. In particular, the parameters used for plotting the resulting resistance curves are somewhat different from the correct values for GeTe. Yet, a qualitative similarity[3] between the results obtained from our

[2] We are using $\tau = 40\,\mathrm{s}$ here because this value is in good agreement with the observed relaxation-time constant in several different systems. This number is slightly less than the τ observed in GeTe (as evident from Fig. 5.4), yet it allows a more direct comparison with the non-equilibrium magnetoresponse in other materials (including $\mathrm{LaAlO_3/SrTiO_3}$ and $\mathrm{Au}_x\mathrm{Ge}_{1-x}$).

[3] We note that traces for up and down sweeps are interchanged, i.e. the down sweep has a higher resistivity for $B > 0$ in our model, whereas this has a lower resistivity in the experiments with GeTe. Again, this is unsurprising, as it depends on the conductivities per carrier of the two Rashba bands. In classic transport theory, our account would be incomplete, as we employed the effective masses of the free Rashba system, whereas the



Reprinted with permission from Ref. [100], V. Narayan et al., Phys. Rev. B 100, 024504 (2019). © *2019 American Physical Society.*

Fig. 5.6 Model prediction of **(a)** non-equilibrium Fermi energies and **(b)** the resulting resistance. The parameters used are $\tau = 40\,\mathrm{s}$, $\dot{B} = 1\,\mathrm{T\,h^{-1}}$, $\varepsilon_F = 0.075\,\mathrm{meV}$, and $E_R/\varepsilon_F = 0.01$.

model and the graphs reported on the normal state non-equilibrium magnetoresistance traces in GeTe suggest that our theoretical framework could in principle be seen as the source for the observed effect.

While we do not aim to extend our framework to also model the non-equilibrium superconducting behaviour of the samples reported in Sec. 5.1, it is worth mentioning that the Fermi energy dynamics suggest a dynamical change of the density of states, which in turn could lead to a reduction of the superconducting transition temperature for a continuously swept magnetic field. Therefore, one could also argue that the developed model can act as a source for the non-equilibrium superconducting transport anomalies reported for the GeTe and $LaAlO_3/SrTiO_3$ samples, yet identifying more clearly how this manifests is beyond the scope of this work.

energy dispersion of GeTe in fact holds rather different effective masses, including a flat-band region just above the nodal point. Moreover, as stated above, the underlying transport theory should take into account quantum corrections.

5.4 Conclusions

In summary, we have presented a novel non-equilibrium effect with ultra-slow relaxation timescales that manifests in the magnetoresponse properties of at least two separate 2D Rashba materials. Both systems feature a normal state and superconducting non-equilibrium effect, which both cannot be explained based on conventional frameworks such as ferromagnetic hysteresis, magnetocaloric effects, trapped flux, or nuclear spin dynamics. We have motivated our assumption that the effect arises due to non-equilibrium carrier distributions and created a theoretical framework based on this assumption that assumes a long relaxation-time constant for the chirality of carriers. Based on this framework we model the dynamical Fermi energies and, by assuming plain classical transport theory, the dynamical resistivity in the normal state. These results are rather qualitative and should not be taken as a quantitative confirmation, yet they are in agreement with experimental findings and suggest that our proposed model could be an accurate description. However, the relaxation-time constant for the chirality number C, which has been put in ‘by hand’ in our model has not been justified quantitatively. This is however of great importance given the significant deviation of this number from typical relaxation-time constants in electronic systems. We therefore conclude the study of our proposed model in this chapter, and we continue with the theoretical study of chirality relaxation timescales in the next chapter, which will serve to assess the relevance of the model put forward here.

Chapter 6

Equilibration studies of chirality in strongly Rashba-coupled systems

In this chapter, we study the scattering mechanisms that relax non-equilibrium chirality distributions of charge carriers in Rashba systems and the relaxation timescales associated with these mechanisms. With chirality being a combination of both spin and momentum locked together, we analyse how, at sufficiently low temperatures, this can result in a protection of chirality from phonon scattering events that in non-Rashba materials cause fast relaxation into equilibrium. The remaining dominant relaxation mechanism, which is the inter-carrier Coulomb interaction, is then studied, and the degree to which the helical spin structure serves to weaken scattering via the inter-carrier interaction is assessed.

The focus in this chapter is on the general relaxation mechanisms of chirality in Rashba systems, however we analyse the implications of our findings in the ferroelectric bulk-Rashba semiconductor GeTe, which ranks amongst the systems with the highest observed Rashba coupling and Rashba momentum split [110], and for which we present novel experimental findings on its non-equilibrium mangetoresistance properties in the previous chapter. At several points throughout this work, we therefore derive results and expressions for a general Rashba system and then evaluate these expressions for the GeTe system.

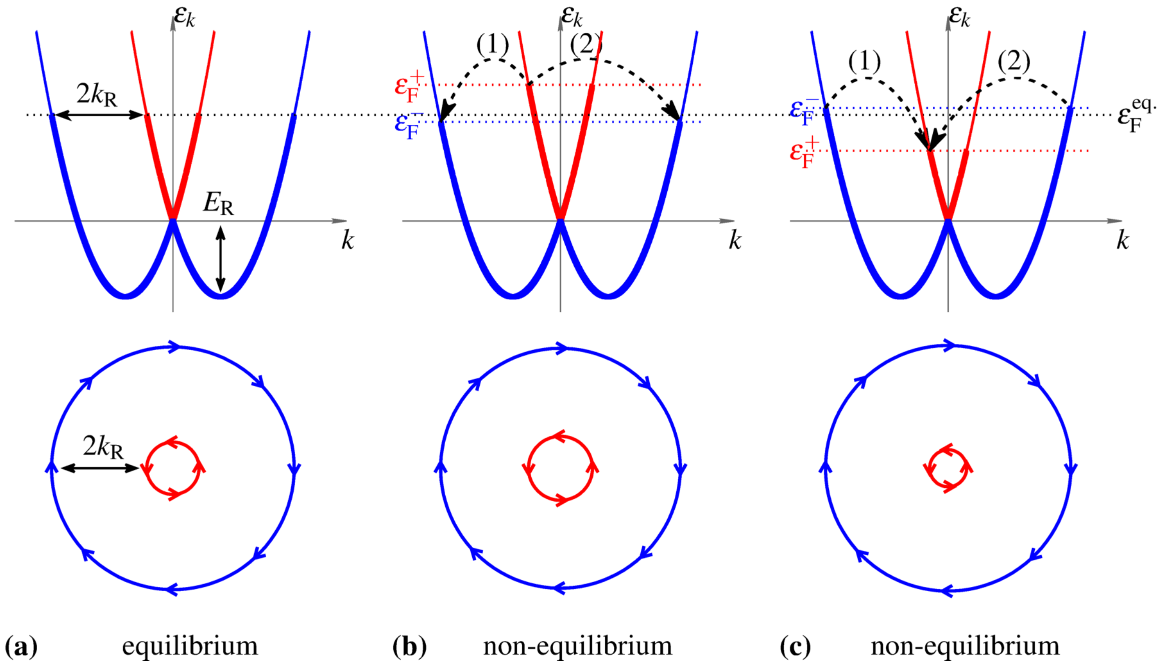
The results are summarised at the end of this chapter, yet a brief summary is presented here. We find that, at sufficiently low temperature, inter-Rashba band transitions become suppressed due to the combined effect of the Rashba momentum split and the chiral spin texture of a Rashba system, which we argue occurs due to a mismatch of the thermal phonon and the Rashba energy-momentum scales at low temperatures. Specifically, we show that momentum exchange between carriers and the phonon bath is effectively absent at temperatures where the momentum of thermal phonons is less than twice the Rashba momentum. This allows us to identify inter-carrier scattering as the dominant process by which non-equilibrium chirality distributions relax. We show that the magnitude of inter-carrier scattering is influenced by the opposing spin structure of the Rashba bands and derive a closed-form expression for the inter-band relaxation timescale associated with inter-carrier Coulomb scattering. We develop a general framework and assess its implications for GeTe, a bulk-Rashba semiconductor with a strong Rashba momentum split, which yields relaxation timescales that are both much longer than conventional electronic relaxation timescales but are at the same time much shorter than the ones assumed in the models employed in the previous chapter.

The organisation of this chapter is as follows. In Sec. 6.1, we provide an introduction to the framework used. Next in Sec. 6.2, we study the relaxation of this non-equilibrium chirality distribution via phonon scattering, and, in Sec. 6.3, we deal with relaxation through the Coulomb interaction. Finally, we conclude in Sec. 6.4. Appendices B.2 and B.3 provide, respectively, explicit calculations of maximum allowed non-equilibrium carrier occupations and the inter-carrier relaxation-time constant, while App. B.4 explains the role of spin-flips in the context of this work.

6.1 Theoretical framework

In this section we define the framework that this study is based on. Again we start from the Rashba band dispersion presented in Sec. 4.1. We investigate the relaxation of this system in a non-equilibrium state depicted in Fig. 6.1, which is very similar to Fig. 5.5 from the previous chapter but without any magnetic fields (we do in fact consider magnetic fields briefly later in this chapter but our primary focus here is on relaxation in the free system).

Fig. 6.1(b) and Fig. 6.1(c) show, respectively, charge carrier configurations in which the Fermi energy of the upper (lower) Rashba band, ε_F^+ (ε_F^-), is higher (lower) than in equilibrium and vice versa. These states are clearly non-equilibrium states and must equilibrate on a finite period of time. As discussed in the previous chapter, the relaxation of such non-equilibrium carrier populations back to equilibrium in any ordinary electronic system can be expected to happen on very short timescales, typically on the order of picoseconds [86]. In this work however, we provide evidence that the mechanisms relaxing this specific non-equilibrium distribution (i.e. chirality imbalance in strongly Rashba-coupled systems at low temperature) are suppressed, which results in much longer relaxation timescales. We note that such carrier populations could either be realised experimentally by injecting spin-polarised currents or by employing strong external magnetic fields. The latter claim is further quantified in Sec. 6.2, and experimental signs for potential realisations of it are in fact discussed in the previous chapter.


ε_k
$2k_\mathrm{R}$
E_R
k
(1)
(2)
ε_F^+
ε_F^-
$\varepsilon_\mathrm{F}^\mathrm{eq.}$
$2k_\mathrm{R}$

(a) equilibrium **(b)** non-equilibrium **(c)** non-equilibrium

Reprinted with permission from Ref. [143], P.C. Verpoort et al., J. Phys.: Condens. Matter 32, 355704 (2020). *Licensed under Creative Commons Attribution 4.0.*

Fig. 6.1 Equilibrium and non-equilibrium chirality distributions of charge carriers in a Rashba system. Top: energy dispersion with bold lines representing occupied states. Bottom: chiral spin structure at Fermi surface with arrows indicating spin directions. **(a)** Equilibrium occupation of the Rashba dispersion. **(b)** Non-equilibrium occupation of the Rashba dispersion. The Fermi energy ε_F^+ (ε_F^-) of the upper (lower) Rashba band is higher (lower) than in equilibrium. In this example, the number of particles in the upper Rashba band is set to be 50% higher than in equilibrium. This number is quite high but was chosen for illustrative purposes. Equilibrium is restored via inter-band transitions, which must involve a spin-flip or a reversal of momentum (labelled 1 and 2, respectively). **(c)** Vice versa of (b) with the number of particles in the upper band being 50% less than in equilibrium.

Relaxation requires scattering events that induce transitions between the Rashba bands. The main process by which carriers can transition between bands is scattering off phonons as these can readily impart very large momentum and affect large-angle and backscattering effects. In addition to phonon scattering, momentum can also be imparted via Coulomb scattering with other carriers (so-called inter-carrier or carrier-carrier scattering) and scattering off charged impurities (so-called carrier-impurity scattering). The present work discusses two key observations for such processes based on the energy-momentum conservation for phonons and the nature of Coulomb scattering. In a system with strong Rashba coupling, the value of k_R can be much larger than the thermal phonon momentum scale at low temperature T. Furthermore, scattering with charged impurities and inter-carrier Coulomb scattering is dominated by the transfer of small momenta, which for inter-band transitions is incompatible with the Rashba coupling because the spinor overlap vanishes for transitions with the smallest possible momentum transfer. These two points are discussed in greater detail in Secs. 6.2 and 6.3.

6.2 Phonon scattering

Charge carriers can exchange momentum with the solid either by emitting phonons or by absorbing or scattering with thermally excited phonons. These processes must obey energy-momentum conservation and therefore we have the following relations for the initial and final momenta, $\mathbf{k}$ and $\mathbf{k}' = \mathbf{k} + \mathbf{q}$, and corresponding energies, $\varepsilon_{\mathbf{k}}$ and $\varepsilon_{\mathbf{k}'}$:

$$\begin{array}{llcl} \text{Emission} : & \varepsilon_{\mathbf{k}} & = & \varepsilon_{\mathbf{k}'} + \hbar\omega_{-\mathbf{q}}\,, \\ \text{Absorption} : & \varepsilon_{\mathbf{k}} + \hbar\omega_{\mathbf{q}} & = & \varepsilon_{\mathbf{k}'}\,, \\ \text{Scattering} : & \varepsilon_{\mathbf{k}} + \hbar\omega_{\mathbf{p}} & = & \varepsilon_{\mathbf{k}'} + \hbar\omega_{\mathbf{p}-\mathbf{q}}\,. \end{array}$$

Here $\hbar\omega_{\mathbf{q}}$ is the energy of a phonon with momentum $\mathbf{q}$, and $\mathbf{p}$ is the momentum of a thermally excited phonon. We assume the carrier initially to be in the band with higher Fermi energy (the inner one in Fig. 6.1(b) and the outer one in Fig. 6.1(c)), and we define $\Delta\varepsilon = \varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}'}$ to be the energy difference between initial and final states. We now combine energy conservation, Pauli exclusion, and the phonon dispersion to construct our argument for the suppression of these relaxation events in the case where T is low such that,

$$k_{\mathrm{B}}T \ll \hbar\omega_{2k_{\mathrm{R}}}. \tag{6.1}$$

Absorption: for an inter-band transition, we require that $q \geq 2k_{\mathrm{R}}$. However at temperatures sufficiently low such that condition Eq. (6.1) holds, such phonons are not thermally excited, thereby disallowing the absorption process completely.

Emission: since for an allowed process $q \geq 2k_{\mathrm{R}}$, we find $\hbar\omega_{2k_{\mathrm{R}}} \leq \hbar\omega_{-\mathbf{q}} = \varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}'} = \Delta\varepsilon$. Consequently, the Rashba momentum scale imposes an upper bound on the energy difference below which inter-band transitions are not allowed:

$$\Delta\varepsilon < \hbar\omega_{2k_{\mathrm{R}}}\,. \tag{6.2}$$

Therefore, the emission is prohibited as long as the energy difference $\Delta\varepsilon$ between any occupied state in the Rashba band with higher Fermi energy and any empty state in the band with lower Fermi energy obeys above condition. This imposes a maximum detuning of the Fermi energies below which equilibration of the non-equilibrium state depicted in Sec. 6.1

will be suppressed. We will discuss the relative strengths of the energy scales of $\Delta\varepsilon$ and $\hbar\omega_{2k_\mathrm{R}}$ in typical Rashba materials at the end of this section. Because this means that there are no unoccupied states to scatter into as the outcome of a phonon emission, one can think of this process as being Pauli-blocked by the occupied low-lying carrier states. Note that all thermally activated excitations will be too small to disobey this condition as long as Eq. (6.1) holds.

Assuming an excess occupation in the Rashba bands of $n^{\pm} = n^{\pm,\mathrm{eq.}}(1 \pm \delta)$ with small δ, where $n^{\pm,\mathrm{eq.}}$ are the equilibrium carrier densities, we find that the condition in Eq. (6.2) is satisfied when approximately

$$\delta < \frac{1}{2}\frac{\hbar\omega_{2k_\mathrm{R}}}{\varepsilon_\mathrm{F}^{\mathrm{eq.}} + E_\mathrm{R}}\,. \tag{6.3}$$

The derivation of this and the exact expression are reported in App. B.2.

Scattering: depending on the relative angle between $\mathbf{q}$ and $\mathbf{p}$, it is $\hbar c_\mathrm{ph}(q-2p) \leq \hbar\omega_{\mathbf{p}-\mathbf{q}} - \hbar\omega_{\mathbf{p}} = \Delta\varepsilon \leq \hbar c_\mathrm{ph} q$, where we have assumed a linear acoustic phonon dispersion with speed of sound c_ph and that $p < q$. Consequently, the modified condition under which scattering is prohibited becomes

$$\Delta\varepsilon < \hbar\omega_{2(k_\mathrm{R}-p)}\,. \tag{6.4}$$

When Eq. (6.1) holds, it is $p \ll k_\mathrm{R}$. Therefore, this condition is almost equivalent to the one for emission. Also note that the above only holds for the case where $\mathbf{p}$ and $\mathbf{q}$ are antiparallel, hence the likelihood of such an event is already diminished in the first place.

It is easy to see how higher-order processes that are built up of several absorption and scattering events could eventually change the momentum of a carrier sufficiently to induce a transition into the other Rashba band while having no strong constraints on the energy difference between initial and final state. The contribution to equilibration can however be expected to be weak, not only because it is a higher-order process but also because the phase space for such a process is small (the relative angles of phonon momenta have to be aligned in a particular way).

Furthermore, it is worth noting that our argument prohibits scattering events independently of whether they conserve or flip the spin of the charge carrier, as our discussion is purely based on energy-momentum conservations.

We now assess the degree to which these effects are present in physical systems. There are two relevant conditions to check, namely whether both $k_\mathrm{B}T$ and $\Delta\varepsilon$ are sufficiently less than $\hbar\omega_{2k_\mathrm{R}}$. Simply speaking, the first condition ensures that there are no thermally excited phonons to absorb for an inter-band transition, whereas the second ensures that the energy detuning is not so big that inter-band transitions can be induced by emission of phonons. For concreteness, we consider GeTe, a system that is known to have giant Rashba coupling, in which $k_\mathrm{R} = 0.19\,\text{Å}^{-1}$ [82]. Assuming a typical value[1] of $c_\mathrm{ph} = 3 \times 10^3\,\mathrm{m\,s^{-1}}$ and $T = 1\,\mathrm{K}$, we find that $k_\mathrm{B}T = 0.09\,\mathrm{meV} \ll \hbar\omega_{2k_\mathrm{R}} = 2\hbar c_\mathrm{ph} k_\mathrm{R} = 7.5\,\mathrm{meV}$. To understand the implications of this on the relaxation timescales, note that the likelihood of such a relaxation event to take place is suppressed exponentially by the a Boltzmann factor of $\exp(-2\hbar c_\mathrm{ph}k_\mathrm{R}/k_\mathrm{B}T) \approx \exp(-87) \approx 10^{-38}$, consequently resulting in effectively a complete suppression of any such phonon-induced relaxation events. Furthermore, the condition in Eq. (6.3) has to be satisfied, which yields $\delta < 1.6\%$ for GeTe with doping of $\varepsilon_\mathrm{F}^\mathrm{eq.} = E_\mathrm{R}/2$ (i.e. close to but above the nodal crossing point, which is realistic for Ge-vacancy doping). Thus, the non-equilibrium state is protected from carrier-phonon relaxation as long as the population imbalance is not more than a few percent. Note that this number can change drastically depending on the Rashba coupling α, the effective mass m, and the speed of sound c_ph.

Before concluding this section, we use the derived expression for the Fermi level detuning in App. B.2 and the upper limit δ found in this section to determine what magnetic field strength would be required to create carrier distributions with such chirality imbalances. Using

$$g\mu_\mathrm{B}B = \Delta\varepsilon_\mathrm{F} = 2\delta(\varepsilon_\mathrm{F} + E_\mathrm{R}) \tag{6.5}$$

as well as $g = 2$, $\delta = 1.6\%$, and $\varepsilon_\mathrm{F} = 0.5E_\mathrm{R}$, we find that $B \approx 9.4\,\mathrm{T}$. The absence of phonon relaxation and the resulting long relaxation times will manifest for carrier imbalances created by magnetic fields below this value, whereas for larger fields the phonon relaxation may no longer be Pauli blocked.

In summary, we have explained how in a Rashba system phonon-induced inter-band transitions of charge carriers are ineffective to relax a detuning of the Fermi level if the detuning is sufficiently small and the temperature is low, both compared to the energy scale $\hbar\omega_{2k_\mathrm{R}}$, which is imposed by the Rashba momentum and the phonon dispersion. We have reasoned that this occurs because phonons are not available for absorption, the emission is Pauli blocked, and higher-order scattering will even for the best possible alignment of phonon momenta have only a comparatively small effect.

[1] This is an estimate that is in good agreement with the c_ph of a vast range of crystals. The value of c_ph in GeTe is unknown, and we therefore use this value for convenience. We note that we do not expect the true value to differ by much. Therefore, our results should not be strongly affected by this approximation.

6.3 Inter-carrier and carrier-impurity scattering

We continue by examining the role of scattering via the Coulomb interaction of charged carriers with each other (inter-carrier scattering) and with charged localised impurities (carrier-impurity scattering). In contrast to the phonon scattering case, here both the Rashba split $2k_\mathrm{R}$ and the opposing helical spin structure at the Fermi surfaces (see Fig. 6.1) play a role in the suppression of the transfer of small momenta. It is worth mentioning that there are no restrictions on these processes *within* a band, but this is irrelevant towards inducing inter-band transitions.

As indicated in Fig. 6.1(b) and Fig. 6.1(c), inter-band transitions can broadly be split up into two types: (1) small-angle scattering events that change the momentum only marginally but have to flip the spin and (2) large-angle scattering that reverses the momentum and leaves the spin unchanged. We show that (1) is strongly suppressed because of the vanishing spinor overlap, which leaves processes of type (2) with large momentum transfer as the dominant mode of relaxation. Coulomb scattering however is dominated by the transfer of small momenta q because the Fourier transform of the Coulomb potential is strong at $q \approx 0$ and because (in the case of inter-carrier scattering) energy conservation is always satisfied when $q = 0$, resulting in a logarithmically divergent phase space. This incompatibility between the nature of the Coulomb interaction and the helical spin alignment of the Rashba energy dispersion is the reason why the Coulomb scattering is also suppressed. However, we also show in the following that this effect is much less pronounced than in the phonon case presented in the previous section.

The entirety of all scattering events includes cases in between (1) and (2), and the aim of the following analysis is to account for this. We derive expressions for the relaxation timescale of carrier-impurity and carrier-carrier scattering, which allows us to show how the above-mentioned arguments manifest quantitatively. Furthermore, we explicitly calculate the timescale for inter-carrier scattering for the case of GeTe. This section only reports the results, whereas detailed calculations can be found in App. B.3. We note that the calculations in this section exclude processes involving spin-flips, whose role we discuss in App. B.4.

Our analysis will be based on Boltzmann transport theory introduced in Sec. 4.2, and we want to study the time dependence of the distribution function $f^{\pm}_{\mathbf{k}_1}$, where the $+$ $(-)$ superscript indicates the upper (lower) Rashba band index. We neglect external fields and temperature gradients, which allows us to reduce the Boltzmann equation to

$$\frac{\partial f^{\pm}_{\mathbf{k}_1}}{\partial t} = I_\mathrm{ci}[f^{\pm}_{\mathbf{k}_1}] + I_\mathrm{cc}[f^{\pm}_{\mathbf{k}_1}] , \tag{6.6}$$

where the indices ci and cc refer to the carrier-impurity and carrier-carrier contributions of the scattering integral, respectively. These are given by

$$
\begin{aligned}
I_{\mathrm{ci}} = \sum_{\mathbf{k}_2} \big(& w^{\mathrm{car-imp}}_{(\mathbf{k}_2\mp)\to(\mathbf{k}_1\pm)}\, f^{\mp}_{\mathbf{k}_2}[1-f^{\pm}_{\mathbf{k}_1}] \\
& - w^{\mathrm{car-imp}}_{(\mathbf{k}_1\pm)\to(\mathbf{k}_2\mp)}\, f^{\pm}_{\mathbf{k}_1}[1-f^{\mp}_{\mathbf{k}_2}]\big)
\end{aligned} \tag{6.7}
$$

and

$$
\begin{aligned}
I_{\mathrm{cc}} = \sum_{\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} \big(& w^{\mathrm{car-car}}_{(\mathbf{k}_3\mathbf{k}_4\mp)\to(\mathbf{k}_1\mathbf{k}_2\pm)}\, f^{\mp}_{\mathbf{k}_3} f^{\mp}_{\mathbf{k}_4}[1-f^{\pm}_{\mathbf{k}_1}][1-f^{\pm}_{\mathbf{k}_2}] \\
& - w^{\mathrm{car-car}}_{(\mathbf{k}_1\mathbf{k}_2\pm)\to(\mathbf{k}_3\mathbf{k}_4\mp)}\, f^{\pm}_{\mathbf{k}_1} f^{\pm}_{\mathbf{k}_2}[1-f^{\mp}_{\mathbf{k}_3}][1-f^{\mp}_{\mathbf{k}_4}]\big)\,.
\end{aligned} \tag{6.8}
$$

We have neglected all terms that conserve the Rashba band index of each particle or induce an exchange of particles between the bands, as these will not lead to a decay of the carrier imbalance between the two Rashba bands. We will make up for this by assuming that *intra*-band scattering events are so quick that they will relax each individual band into *local* thermal equilibrium on a timescale that is immediate compared to the *inter*-band processes.

We calculate the probability amplitudes $w_{(\mathbf{k}_3\mathbf{k}_4\mp)\to(\mathbf{k}_1\mathbf{k}_2\pm)}$ using Fermi's Golden Rule,

$$
\begin{aligned}
w_{(\mathbf{k}_3\mathbf{k}_4\mp)\to(\mathbf{k}_1\mathbf{k}_2\pm)} = & \frac{2\pi}{\hbar}\, |\langle\Psi_{\mathrm{final}}|\,U\,|\Psi_{\mathrm{init}}\rangle|^2 \\
& \times\ \delta(\varepsilon^{\pm}_{\mathbf{k}_1}+\varepsilon^{\pm}_{\mathbf{k}_2}-\varepsilon^{\mp}_{\mathbf{k}_3}-\varepsilon^{\mp}_{\mathbf{k}_4})\,,
\end{aligned} \tag{6.9}
$$

for which we need to obtain the matrix element corresponding to the relevant transition. The matrix element will consist of two parts: 1) the Fourier transform of the Coulomb potential, which arises from its expectation value for the incoming and outgoing plane waves and 2) the spinor overlap between initial and final states. We shall use the 2D Fourier transform of the screened 3D Coulomb potential for our calculations (where the static screening is obtained using the random-phase approximation) [119], which is of the form

$$
U_{\mathbf{p}} = U_p = \frac{2\pi e_0^2}{p+k_{\mathrm{S}}} \tag{6.10}
$$

with $e_0^2 = e^2/4\pi\kappa\varepsilon_0$, where κ is the effective background lattice dielectric constant, k_{S} is the Thomas-Fermi screening momentum, and $\mathbf{p} = \mathbf{k}-\mathbf{k}'$, where $\mathbf{k}$ and $\mathbf{k}'$ are the incoming and outgoing plane waves. Furthermore, to determine the spinor overlap, we write each single-particle state as a superposition of the Pauli matrix σ_Z eigenstates [129] as

$$|\mathbf{k},+\rangle = \frac{1}{\sqrt{2}}\left(|\mathbf{k},\uparrow\rangle - \mathrm{i}e^{\mathrm{i}\theta_{\mathbf{k}}}\,|\mathbf{k},\downarrow\rangle\right), \tag{6.11}$$

$$|\mathbf{k},-\rangle = \frac{1}{\sqrt{2}}\left(-\mathrm{i}e^{-\mathrm{i}\theta_{\mathbf{k}}}\,|\mathbf{k},\uparrow\rangle + |\mathbf{k},\downarrow\rangle\right), \tag{6.12}$$

where $\theta_{\mathbf{k}}$ is defined such that $\mathbf{k} = (k_x,k_y)^{\mathsf{T}} = |\mathbf{k}|\;(\cos\theta_{\mathbf{k}},\,\sin\theta_{\mathbf{k}})^{\mathsf{T}}$. This can then be used to evaluate the overlap between states from different bands, $\langle\mathbf{k}',+|\,U\,|\mathbf{k},-\rangle$, as

$$\mathrm{e}^{-\frac{\mathrm{i}}{2}\left(\theta_{\mathbf{k}}+\theta_{\mathbf{k}'}\right)}\,\sin\left(\theta_{\mathbf{k}}-\theta_{\mathbf{k}'}/2\right)\,\left\langle\mathbf{k}'\right|U\left|\mathbf{k}\right\rangle. \tag{6.13}$$

Following a well-known approach by Yafet [158], which derives an expression for the spin relaxation time from phonon-assisted spin-flip processes, we start from two Rashba bands, induce a small imbalance of the chemical potential and perturb to first order in the detuning to find an expression for the relaxation-time constant for carrier-impurity scattering, τ_{ci}, and carrier-carrier scattering, τ_{cc}, respectively.

Our calculation yields the following expression for the relaxation-time constant of the carrier-impurity processes.

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{1}{\tau_{\mathrm{ci},0}}\frac{\pi}{4}, \tag{6.14}$$

where $\tau_{\mathrm{ci},0}^{-1} = 8\pi m n_{\mathrm{i}}\, e_0^4/\hbar^3 (k_{\mathrm{F}}^0)^2$, n_{i} is the impurity density (per area), and k_{F}^0 is the average equilibrium Fermi momentum of the system. We compare this result to the case where we omit an essential feature of the Rashba system, namely the helical alignment of spin eigenstate at the Fermi surface, whose overlap is given by the first two factors in Eq. (6.13). This yields

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{1}{\tau_{\mathrm{ci},0}}\frac{k_{\mathrm{F}}^0}{k_{\mathrm{S}}}, \tag{6.15}$$

which is enhanced over the result in Eq. (6.14) by a factor of $\frac{k_{\mathrm{F}}^0}{k_{\mathrm{S}}}$. Since typically the Fermi momentum is much larger than the screening momentum scale, we see how the Rashba dispersion serves to enhance the screening of the Coulomb interaction. Using standard Lindhard theory, we estimate $\frac{k_{\mathrm{F}}^0}{k_{\mathrm{S}}} \approx 23.7$ for GeTe. This result is worth noting but equally not too relevant for practical applications, where the carrier-impurity scattering is mainly influenced by the impurity concentration n_{i} and the suppression by a factor of $\frac{k_{\mathrm{F}}^0}{k_{\mathrm{S}}}$ will not be as big as in the case of electron-phonon scattering described above.

Furthermore, we derive the following expression for the carrier-carrier Coulomb scattering relaxation-time constant,

$$\frac{1}{\tau_{\mathrm{cc}}} = \frac{1}{\tau_{\mathrm{cc},0}} \frac{(k_{\mathrm{B}}T)^2}{(\mu - E_{\mathrm{R}})^2} \times \rho\left(\frac{E_{\mathrm{R}}}{\mu}, \frac{k_{\mathrm{B}}T}{\mu - E_{\mathrm{R}}}\right), \tag{6.16}$$

with

$$\rho\,(x,y) = \frac{\pi^2}{6}\left(1 - x(1 - \log(x)) - (1-x)\log\left(\frac{\pi^2}{6}y\right)\right), \tag{6.17}$$

where $\tau_{cc,0}^{-1} = \frac{(2\pi e_0^2)^2}{2\pi\hbar}\frac{2m}{\hbar^2}$.

This result is of a form similar to the scattering lifetime of a quasi particle subject to inter-carrier Coulomb interaction reported by Zheng and Das Sarma [162]. The expression for ρ contains a logarithmically divergent part and a constant part, where the latter occurs due to the regularising effect of the opposing helical spin structure (which however only affects one part of the divergent phase space).

Using $T = 100\,\mathrm{mK}$ as well as $\kappa = 10$, $\mu - E_{\mathrm{R}} = 0.1\,\mathrm{eV}$, and $\frac{E_{\mathrm{R}}}{\mu} \approx 0.5$ (which we assume for 32 nm thick α-GeTe [82, 99, 140]), we find

$$\tau_{\mathrm{cc}} \approx 1\,\mu\mathrm{s}\,. \tag{6.18}$$

This result is impressive in that the lifetime is significantly enhanced over the picosecond lifetime that is commonly observed in electronic systems [86]. Normally, relaxation timescales on the order of microseconds are only observed for carrier-phonon processes at extremely low temperature [68]. However, we also note that the suppression is not of the exponential form found for the phonon case and hence not as dramatic. Furthermore, we note that the chiral alignment of spin eigenstates on the Fermi surface only serves to suppress one of the two logarithmic divergences at $q = 2k_{\mathrm{R}}$ of the form $\log({}^{\mu - E_{\mathrm{R}}}/_{k_{\mathrm{B}}T})$ and does not affect the same divergence occurring at $q = k_{\mathrm{F}}^{+} + k_{\mathrm{F}}^{-}$. It would be interesting to see if more complex band structures involving more intricate spin alignments could result in even stronger suppression and even longer relaxation timescales.

We note that, while our calculation is sufficiently accurate to approximate the order of magnitude of the relaxation-time constant, an estimate of the time constants for real materials necessitates the consideration of exact band structures as well as, possibly, dynamic screening and exchange interaction effects.

6.4 Discussion and conclusions

In summary, this chapter studies the relaxation of chirality imbalances in Rashba systems, and the relaxation processes and timescales associated with those states. It is shown that phonon-mediated inter-band transitions in Rashba systems are effectively absent when T is low and the occupation number detuning δ is below a certain threshold. This happens due to an energy-momentum mismatch between the electronic and phonon dispersion, and consequently phonons do not contribute to relaxation of carrier chirality. We identify the inter-carrier scattering mediated by the Coulomb interaction as the resulting dominant relaxation mechanism (in a pure sample) and further analyse it. We find that this is also weakened due to the chiral spin structure at the Fermi level, and consequently the relaxation time arising from this is much longer than what is commonly expected from inter-carrier interaction. We estimate the relaxation timescales for the inter-carrier scattering for a typical strongly Rashba-coupled system (GeTe in our example) at low temperatures of $T \approx 100\,\mathrm{mK}$ to be on the order of $\tau \approx 1\,\mu\mathrm{s}$.

This result is on the one hand surprising because 1 µs is certainly beyond what is commonly observed as a relaxation timescale in most electronic systems, on the other hand it is much shorter than the experimentally observed timescale of several minutes reported for the novel non-equilibrium effect in GeTe in the previous chapter. This allows us to conclude that, despite the intriguing idea of modelling the dynamical magentoresistance curves presented in the previous chapter, chirality imbalances are unlikely to be the sole source of the observed effects since such imbalances would decay on much faster timescales.

Nonetheless, it might be interesting to consider whether the validity of the theoretical predictions reported in this chapter could be tested through experimental investigations. We argue that the studied non-equilibrium chirality populations could be realised in experiments in a number of ways, such as through the application of magnetic fields as well as by injecting spin-polarised currents that dominantly occupy one of the two Rashba bands, although developing these ideas further is beyond the scope of this work.

We note that the results we obtain for the timescales mainly rely on the large momentum split that strong Rashba coupling induces between the Fermi surfaces of the bands of the two carrier species as well as on the fact that carriers from different Rashba bands with lowest momentum separation have orthogonal spin states (which is always true for systems with a spherical Fermi surface). As such, it is possible to generalise the results from this study to systems with different energy dispersion, as long as those two main ingredients are retained.

More generally, a remarkable finding of our study is that, due to the absence of phonons, the dominant mode of inter-band relaxation is carrier-carrier Coulomb scattering, whereas conventionally this is mediated by phonons.

Chapter 7

Conclusion

This thesis studies the equilibration properties of two distinct systems, starting with the minimisation dynamics of an SGD optimiser in the LFL of a DNN, followed by the relaxation of non-equilibrium chirality distributions of charge carriers in low-temperature strongly Rashba-coupled solid-state systems.

In Part I of this thesis, we advance the understanding of the training process of DNNs with SGD by gaining a better understanding of the underlying LFL structure. As outlined in Chap. 2, the aim of our investigation is to shine light into the the unusual effectiveness of SGD optimisation, which is commonly observed in deep learning. While extreme values of the learning rate result in either too small or too large noise to facilitate optimisation, the parameters of SGD can in most cases be easily tuned to promote a global downhill trajectory. This crucial observation is best summarised by Fig. 2.2. This effectiveness of deep learning is unusual given the complexity of the optimisation problem involved in minimising the loss function, which is high-dimensional and non-convex. It has empirically been observed by the ML community that this principle is almost universally applicable. Therefore, a clear understanding of this conundrum is eminent and holds the potential to design even more effective ML methods.

Following this introduction, Chap. 3 subsequently analyses the LFLs of a range of DNN applications using different datasets (LJAT19, OPTDIG, and WINE). For LJAT19 we assess the dependency of the LFL structure on the network deepness ($H = 1,2,3$) and amount of training data ($N_{\text{data}} = 100, 1000, 2000, 10000, 100000$). Our main finding is that for all datasets and hyperparameters, the local optima in the LFL are connected by low barriers, which is apparent from the disconnectivity graphs in Figs. 3.2, 3.5, and 3.6. This is true across all datasets and persists for different N_{data} and H. More precisely, for all shallow nets with $H = 1$ and in the deep case $H > 1$ with a sufficiently large N_{data} (i.e. $N_{\text{data}} \gtrsim 2000$), the LFL features a single funnel, where in average the downhill paths from a minimum to the global minimum display small barriers. Only in the deep data-scarce limit, we observe a novel, previously unobserved landscape structure, which comprises many local minima of similar loss values connected by low barriers.

These results are intriguing and lead to the final conclusion that one of the main reasons for the success of SGD optimisation methods in deep learning is the underlying LFL structure of the employed DNN architectures. The geometry that these LFLs display can be described as ‘nearly convex’, which is easy to optimise by methods as simple as SGD. Moreover, we analyse the correlation between train loss, test loss, and minima geometry, where the latter is quantified in terms of the Hessian eigenvalues and their log product. The results confirm previously reported evidence of the enhanced generalisability of minima that are ‘wide’ and ‘shallow’ [10, 26] but also reveals new results concerning the precise correlation between loss generalisability and minima geometry.

In the future, it would be interesting to extend the studies of LFLs reported in this work to consider other common ML models. This should entail the investigation of bigger systems, given that all architectures and datasets used in our work are relatively small compared to those commonly employed in ML applications. While the results presented here show clear trends that we believe will generalise to larger networks, it remains crucial to ascertain the continuation of this trend beyond the small networks considered in this contribution.

Moreover, there are many hyperparameters, such as the regularisation parameter λ or the non-linear activation functions ϕ_l, whose influence onto the LFL structure would be intriguing to investigate. There also exist many other popular ML architectures, such as Convolutional Neural Networks or Recursive Neural Networks, which follow quite similar principles as DNNs and which would lend themselves perfectly as a way to extend the methods used in this work to a broader study of LFLs in ML models.

Importantly, one extension of the work presented here would include a more accurate identification of permutational isomers of minima and TSs in the LFLs of DNNs and avoid falsely identifying newly discovered stationary points as symmetry-related representations of already known ones due to an insufficiently small loss difference tolerance in cases of LFLs with a high density of stationary points. While we do not expect this deficiency to affect the overarching results of this work, obtaining a complete database of minima and TSs for the simple DNN examples considered here would be desirable and should be a central aim for future work.

Furthermore, it would be exciting to explore what other conclusions can be drawn from the resolved LFL structure. For instance, sampling techniques, which rely on the combination of the predictions of several ML models to obtain improved and more robust models, would be interesting to investigate using a LFL analysis. This approach could in turn have the potential to reveal novel ML techniques that stem from advanced landscape exploration tools. Finally, a study of higher-index saddle points, which are stationary points but with a Hessian index higher than 1, could complement our understanding of DNN optimisation, since many commonly employed DNN optimisers in fact converge to such points.

Moving on to Part II of this thesis, Chaps. 4, 5, and 6 study the relaxational dynamics of non-equilibrium charge carrier distributions in strongly Rashba-coupled systems at low temperature. We define the chirality number C of those carrier distributions, i.e. the difference between numbers of carriers with chirality $+$ and $-$, and we investigate the equilibration of this quantity to its equilibrium value $C_{\mathrm{eq.}}$ with time.

Chaps. 5 and 6 present the two main studies of non-equilibrium carrier distributions of this type, best summarised by Figs. 5.5 and 6.1. First, we motivate the existence of such non-equilibrium configurations in Chap. 5 based on experimental evidence found in the magnetotransport properties of low-temperature Rashba systems subject to external magnetic fields, where the most prominent salient features are shown in Figs. 5.3 and 5.4. This is followed by a theoretical study of the dynamical Fermi energies of the two Rashba bands, which is governed by the differential equation Eq. (5.3). Using a fixed relaxation-time constant τ, integrating Eq. (5.3), and plotting its results in Fig. 5.6 allows us to conclude that some of the measurements reported in Chap. 5 can at least partly be modelled using our framework, which therefore can be seen as a potential source of the observed effect. However, uncertainty regarding various aspects of this model persist, in particular with respect to the system-specific properties assumed in the modelling of the resistance graphs, as well as in relation to the role of the quantum corrections that result in the WAL cusp. These remaining inconsistencies result in a cautious interpretation of the findings derived from our model. Moreover, the value of the relaxation-time constant of several seconds employed in the model is unusually long for electronic relaxation timescales and hence requires further justification on a microscopic level that goes beyond the pure qualitative assessment of the experimental observations.

A quantitative assessment of the relaxation timescales is therefore conducted in Chap. 6, where we study the relaxation of the non-equilibrium chirality distributions through phonon and Coulomb-interaction processes. We find that, while phonon-mediated relaxation may be practically absent at low temperatures of $T = 1\,\mathrm{K}$ and for a small Fermi level detuning δ of a few percent in GeTe, the inter-carrier Coulomb scattering persists at moderately strong magnitude, resulting in a relaxation-time constant of $\tau = 1\,\mu\mathrm{s}$. Consequently, this finding casts some doubt on the validity of the model developed in Chap. 5 and suggests that a more thorough analysis of the charge-carrier configuration in GeTe is needed to adequately describe the discovered novel non-equilibrium effect in low-temperature Rashba systems.

The studies presented in the second part of this thesis could be extended in numerous ways in the future. First and foremost, the resolution of the puzzling magnetic-field induced non-equilibrium transport properties in GeTe and other Rashba systems should be revisited, and new frameworks taking into account other sample-related properties should be considered. In addition, experiments further probing the salient features observed in those samples would be extremely insightful. For example, investigating the frequency dependence of the magnetotransport properties beyond the simple AC and DC limits presented in Fig. 5.3 would be of great relevance. Moreover, a careful study of the temperature dependence of the observed relaxation-time constants could provide insights into the underlying relaxation mechanism and hence the nature of the observed non-equilibrium effect.

On the theoretical side, there exist two ways of extending the presented work. On the one hand, it could be fruitful to enhance the current framework and incorporate other features of the studied system, which could result in a suitable model for the reported experimental observations. As such, our current assessment of the relaxation-time constant only considers the relaxation of charge carriers. However, it is evident from the experiments that the system is close to a superconducting instability, which suggests that perhaps the relaxation of Cooper pairs may in fact be a property worth investigating. Moreover, the dispersion relation of holes in the valence band of GeTe features flat-band regions, which are subject to various instabilities. Studying these could reveal alternative mechanisms resulting in the observed long relaxation-time constants in the magnetotransport properties of GeTe and other Rashba materials. Finally, an assessment of the results for the relaxation timescales reported in Chap. 6 independent of the experimental evidence reported in Chap. 5 could be interesting. The predicted relaxation-time constant of $\tau = 1\,\mu\mathrm{s}$ could potentially become observable in experiments, for example in devices that combine strong Rashba coupling and spin-polarised currents.

References

[1] Aarts, E. and Korst, J. (1988). *Simulated Annealing and Boltzmann Machines*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., USA.

[2] Alpaydin, E. and Kaynak, C. (1998). Cascading classifiers. *Kybernetika*, 34(4):369.

[3] Altshuler, B. L., Khmel'nitzkii, D., Larkin, A. I., and Lee, P. A. (1980). Magnetoresistance and Hall effect in a disordered two-dimensional electron gas. *Phys. Rev. B*, 22:5142.

[4] Andrews, B. and Conduit, G. (2020). Absence of diagonal force constants in cubic Coulomb crystals. *Proc. R. Soc. A.*, 476(2244):20200518.

[5] Ashcroft, N. W. and Mermin, N. D. (1976). *Solid State Physics*. Hartcourt College Publishers, Orlando.

[6] Aslamazov, L. G. (1969). Influence of impurities on the existence of an inhomogeneous state in a ferromagnetic superconductor. *Soviet Phys. JETP*, 28(4):773.

[7] Axilrod, B. M. and Teller, E. (1943). Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.*, 11:299.

[8] Backes, D., Huang, D., Mansell, R., Lanius, M., Kampmeier, J., Ritchie, D., Mussler, G., Gumbs, G., Grützmacher, D., and Narayan, V. (2019). Thickness dependence of electron-electron interactions in topological $p-n$ junctions. *Phys. Rev. B*, 99:125139.

[9] Baity-Jesi, M., Sagun, L., Geiger, M., Spigler, S., Arous, G. B., Cammarota, C., LeCun, Y., Wyart, M., and Biroli, G. (2019). Comparing dynamics: deep neural networks versus glassy systems. *J. of Stat. Mech.: Theory and Experiment*, 2019(12):124013.

[10] Baldassi, C., Borgs, C., Chayes, J. T., Ingrosso, A., Lucibello, C., Saglietti, L., and Zecchina, R. (2016). Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes. *Proc. National Academy of Sci. USA*, 113(48):7655.

[11] Ballard, A. J., Das, R., Martiniani, S., Mehta, D., Sagun, L., Stevenson, J. D., and Wales, D. J. (2017). Energy landscapes for machine learning. *Phys. Chem. Chem. Phys.*, 19:12585.

[12] Ballard, A. J., Stevenson, J. D., Das, R., and Wales, D. J. (2016). Energy landscapes for a machine learning application to series data. *J. Chem. Phys.*, 144(12):124119.

[13] Baral, A., Vollmar, S., Kaltenborn, S., and Schneider, H. C. (2016). Re-examination of the Elliott–Yafet spin-relaxation mechanism. *New J. of Phys.*, 18(2):023012.

[14] Baraniuk, R., Donoho, D., and Gavish, M. (2020). The science of deep learning. *Proc. Natl. Acad. Sci. USA*, 117(48):30029–30032.

[15] Becker, O. M. and Karplus, M. (1997). The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.*, 106:1495.

[16] Bercioux, D. and Lucignano, P. (2015). Quantum transport in Rashba spin–orbit materials: a review. *Reports on Prog. in Phys.*, 78(10):106001.

[17] Bihlmayer, G., Rader, O., and Winkler, R. (2015). Focus on the Rashba effect. *New J. of Phys.*, 17(5):050202.

[18] Bishop, C. (1992). Exact Calculation of the Hessian Matrix for the Multilayer Perceptron. *Neural Computation*, 4(4):494.

[19] Bottou, L. (2010). Large-Scale Machine Learning with Stochastic Gradient Descent. In Lechevallier, Y. and Saporta, G., editors, *Proceedings of COMPSTAT'2010*, page 177. Physica-Verlag, Heidelberg.

[20] Brinkman, A., Huijben, M., van Zalk, M., Huijben, J., Zeitler, U., Maan, J. C., van der Wiel, W. G., Rijnders, G., Blank, D. H. A., and Hilgenkamp, H. (2007). Magnetic effects at the interface between non-magnetic oxides. *Nature Materials*, 6:493.

[21] Brooks, B. R., Brooks III, C. L., Mackerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614.

[22] Broyden, C. G. (1970). The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *J. Inst. Math. Appl.*, 6:76.

[23] Bychkov, Y. A. and Rashba, É. I. (1984). Properties of a 2D electron gas with lifted spectral degeneracy. *JETP Lett.*, 39(2):78.

[24] Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005). Finding pathways between distant local minima. *J. Chem. Phys.*, 122:234903.

[25] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688.

[26] Chaudhari, P., Choromanska, A., Soatto, S., LeCun, Y., Baldassi, C., Borgs, C., Chayes, J., Sagun, L., and Zecchina, R. (2019). Entropy-SGD: biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124018.

[27] Chebaro, Y., Ballard, A. J., Chakraborty, D., and Wales, D. J. (2015). Intrinsically Disordered Energy Landscapes. *Scientific Reports*, 5:10386.

[28] Chen, N.-K., Li, X.-B., Bang, J., Wang, X.-P., Han, D., West, D., Zhang, S., and Sun, H.-B. (2018). Directional Forces by Momentumless Excitation and Order-to-Order Transition in Peierls-Distorted Solids: The Case of GeTe. *Phys. Rev. Lett.*, 120:185701.

[29] Chill, S. T., Stevenson, J., Ruehle, V., Shang, C., Xiao, P., Farrell, J. D., Wales, D. J., and Henkelman, G. (2014). Benchmarks for Characterization of Minima, Transition States, and Pathways in Atomic, Molecular, and Condensed Matter Systems. *J. Chem. Theor. Comput.*, 10(12):5476.

[30] Chitturi, S. R., Verpoort, P. C., Lee, A. A., and Wales, D. J. (2020). Perspective: new insights from loss function landscapes of neural networks. *Machine Learning: Science and Technology*, 1(2):023002.

[31] Choromanska, A., Henaff, M., Mathieu, M., Arous, G., and LeCun, Y. (2015). The Loss Surfaces of Multilayer Networks. In *Proceedings of Machine Learning Research*, volume 38, page 192. Machine Learning Research Press.

[32] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (1998). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547.

[33] Ćustić, A., Sokol, V., Punnen, A. P., and Bhattacharya, B. (2017). The bilinear assignment problem: complexity and polynomially solvable special cases. *Mathematical Programming*, 166(1-2):185.

[34] Dann, J., Verpoort, P., Ferreira de Oliveira, J., Rowley, S., Datta, A., Kar-Narayan, S., Ford, C., Conduit, G., and Narayan, V. (2019). Au-Ge Alloys for Wide-Range Low-Temperature On-Chip Thermometry. *Phys. Rev. Applied*, 12:034024.

[35] Daptary, G. N., Kumar, S., Bid, A., Kumar, P., Dogra, A., Budhani, R. C., Kumar, D., Mohanta, N., and Taraphder, A. (2017). Observation of transient superconductivity at the $\mathrm{LaAlO_3/SrTiO_3}$ interface. *Phys. Rev. B*, 95:174502.

[36] Das, R. and Wales, D. J. (2016). Energy landscapes for a machine-learning prediction of patient discharge. *Phys. Rev. E*, 93:063310.

[37] Das, R. and Wales, D. J. (2017). Machine learning prediction for classification of outcomes in local minimisation. *Chem. Phys. Lett.*, 667:158.

[38] Datta, S. and Das, B. (1990). Electronic analog of the electro-optic modulator. *Appl. Phys. Lett.*, 56(7):665.

[39] de Souza, V. K., Stevenson, J. D., Niblett, S. P., Farrell, J. D., and Wales, D. J. (2017). Defining and quantifying frustration in the energy landscape: Applications to atomic and molecular clusters, biomolecules, jammed and glassy systems. *J. Chem. Phys.*, 146(12):124103.

[40] de Souza, V. K. and Wales, D. J. (2008). Energy landscapes for diffusion: Analysis of cage-breaking processes. *J. Chem. Phys.*, 129:164507.

[41] de Souza, V. K. and Wales, D. J. (2009). Connectivity in the potential energy landscape for binary Lennard-Jones systems. *J. Chem. Phys.*, 130:194508.

[42] Deng, L. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, 7(3-4):197–387.

[43] Di Sante, D., Paolo, B., Riccardo, B., and Silvia, P. (2012). Electric Control of the Giant Rashba Effect in Bulk GeTe. *Advanced Materials*, 25(4):509.

[44] Dil, J. H. (2009). Spin and angle resolved photoemission on non-magnetic low-dimensional systems. *J. of Phys.: Condensed Matter*, 21(40):403001.

[45] Draxler, F., Veschgini, K., Salmhofer, M., and Hamprecht, F. A. (2018). Essentially No Barriers in Neural Network Energy Landscape. In *Proceedings of Machine Learning Research*, volume 80, page 1309. Machine Learning Research Press.

[46] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[47] Edwards, A. H., Pineda, A. C., Schultz, P. A., Martin, M. G., Thompson, A. P., and Hjalmarson, H. P. (2005). Theory of persistent, p-type, metallic conduction in c-GeTe. *J. of Phys.: Condensed Matter*, 17(32):L329.

[48] Edwards, A. H., Pineda, A. C., Schultz, P. A., Martin, M. G., Thompson, A. P., Hjalmarson, H. P., and Umrigar, C. J. (2006). Electronic structure of intrinsic defects in crystalline germanium telluride. *Phys. Rev. B*, 73:045210.

[49] Elliott, R. J. (1954). Theory of the Effect of Spin-Orbit Coupling on Magnetic Resonance in Some Semiconductors. *Phys. Rev.*, 96:266.

[50] Fanciulli, M., Lei, T., and Moustakas, T. D. (1993). Conduction-electron spin resonance in zinc-blende GaN thin films. *Phys. Rev. B*, 48:15144.

[51] Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13:317.

[52] Fulde, P. and Ferrell, R. A. (1964). Superconductivity in a Strong Spin-Exchange Field. *Phys. Rev.*, 135:A550.

[53] Gavin, H. (2020). The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems. Available Online from [http://people.duke.edu/ hpgavin/ce281/lm.pdf], accessed Dec 12 2020.

[54] Gibbs, J. W. and Tyndall, J. (1874). *On the equilibrium of heterogeneous substances: first part*. Connecticut Academy of Arts and Sciences, Connecticut.

[55] Giussani, A., Perumal, K., Hanke, M., Rodenbach, P., Riechert, H., and Calarco, R. (2012). On the epitaxy of germanium telluride thin films on silicon substrates. *Phys. Status Solidi B*, 249(10):1939.

[56] Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Math. Comput.*, 24:23.

[57] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29.

[58] Hein, R. A., Gibson, J. W., Mazelsky, R., Miller, R. C., and Hulm, J. K. (1964). Superconductivity in Germanium Telluride. *Phys. Rev. Lett.*, 12:320.

[59] Henkelman, G. and Jónsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, 113:9978.

[60] Henkelman, G., Uberuaga, B. P., and Jónsson, H. (2000). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, 113:9901.

[61] Hikami, S., Larkin, A. I., and Nagaoka, Y. (1980). Spin-Orbit Interaction and Magnetoresistance in the Two Dimensional Random System. *Prog. of Theor. Phys.*, 63(2):707.

[62] Hochreiter, S. and Schmidhuber, J. (1994). Simplifying neural nets by discovering flat minima. In Tesauro, G., Touretzky, D., and Leen, T., editors, *Advances in Neural Information Processing Systems*, volume 7. MIT Press, Cambridge, Massachusetts.

[63] Hochreiter, S. and Schmidhuber, J. (1997). Flat Minima. *Neural Computation*, 9(1):1.

[64] Houzet, M. and Meyer, J. S. (2015). Quasiclassical theory of disordered Rashba superconductors. *Phys. Rev. B*, 92:014509.

[65] Jastrzębski, S., Kenton, Z., Arpit, D., Ballas, N., Fischer, A., Bengio, Y., and Storkey, A. (2017). Three Factors Influencing Minima in SGD. *arXiv e-prints*. arXiv:1711.04623.

[66] Jones, J. E. and Ingham, A. E. (1925). On the calculation of certain crystal potential constants, and on the cubic crystal of least potential energy. *Proc. R. Soc. Lond. A*, 107:636.

[67] Joseph, J. A., Röder, K., Chakraborty, D., Mantell, R. G., and Wales, D. J. (2017). Exploring biomolecular energy landscapes. *Chem. Commun.*, 53:6974.

[68] Karvonen, J. T. and Maasilta, I. J. (2007). Influence of Phonon Dimensionality on Electron Energy Relaxation. *Phys. Rev. Lett.*, 99:145503.

[69] Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, page 1942.

[70] Keskar, N., Nocedal, J., Tang, P., Mudigere, D., and Smelyanskiy, M. (2017). On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. In *5^{th} International Conference on Learning Representations, ICLR 2017*.

[71] Kim, M.-C., Park, H., Son, S., Sim, E., and Burke, K. (2015). Improved DFT Potential Energy Surfaces via Improved Densities. *J. Phys. Chem. Lett.*, 6(19):3802–3807.

[72] Kittel, C. and Kroemer, H. (1980). *Thermal Physics*. W.H. Freeman and Company, New York.

[73] Kolobov, A. V., Tominaga, J., Fons, P., and Uruga, T. (2003). Local structure of crystallized GeTe films. *Applied Phys. Lett.*, 82(3):382.

[74] Krempaský, J., Fanciulli, M., Pilet, N., Minár, J., Khan, W., Muntwiler, M., Bertran, F., Muff, S., Weber, A., Strocov, V., Volobuiev, V., Springholz, G., and Dil, J. (2019). Spin-resolved electronic structure of ferroelectric α-GeTe and multiferroic $Ge_{1-x}Mn_xTe$. *J. of Phys. and Chem. of Solids*, 128:237.

[75] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83.

[76] Larkin, A. I. and Ovchinnikov, I. U. N. (1965). Inhomogeneous state of superconductors (Production of superconducting state in ferromagnet with Fermi surfaces, examining Green function). *Soviet Phys. JETP*, 20(3):762.

[77] LaShell, S., McDougall, B. A., and Jensen, E. (1996). Spin Splitting of an Au(111) Surface State Band Observed with Angle Resolved Photoelectron Spectroscopy. *Phys. Rev. Lett.*, 77:3419.

[78] Lawrence, W. E. and Wilkins, J. W. (1973). Electron-Electron Scattering in the Transport Coefficients of Simple Metals. *Phys. Rev. B*, 7:2317.

[79] Li, H., Xu, Z., Taylor, G., Studer, C., and Goldstein, T. (2018). Visualizing the Loss Landscape of Neural Nets. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., Red Hook, USA.

[80] Li, Z. and Scheraga, H. A. (1987). Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA*, 84:6611.

[81] Li, Z. and Scheraga, H. A. (1988). Structure and free energy of complex thermodynamic systems. *J. Mol. Struct.*, 179:333.

[82] Liebmann, M., Rinaldi, C., Di Sante, D., Kellner, J., Pauly, C., Wang, R. N., Boschker, J. E., Giussani, A., Bertoli, S., Cantoni, M., Baldrati, L., Asa, M., Vobornik, I., Panaccione, G., Marchenko, D., Sánchez-Barriga, J., Rader, O., Calarco, R., Picozzi, S., Bertacco, R., and Morgenstern, M. (2016). Giant Rashba-Type Spin Splitting in Ferroelectric GeTe(111). *Advanced Materials*, 28(3):560.

[83] Liu, Y. Y. F., Andrews, B., and Conduit, G. J. (2019). Direct evaluation of the force constant matrix in quantum Monte Carlo. *J. Chem. Phys.*, 150(3):034104.

[84] Luo, R., Tian, F., Qin, T., Chen, E., and Liu, T.-Y. (2018). Neural architecture optimization. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31, page 7816. Curran Associates, Inc., Red Hook, USA.

[85] Manchon, A., Koo, H. C., Nitta, J., Frolov, S. M., and Duine, R. A. (2015). New perspectives for Rashba spin–orbit coupling. *Nat. Mater.*, 14:871.

[86] Marder, M. P. (2010). *Condensed Matter Physics*. John Wiley & Sons, Inc., Hoboken, New Jersey.

[87] Matsunaga, T., Fons, P., Kolobov, A. V., Tominaga, J., and Yamada, N. (2011). The order-disorder transition in GeTe: Views from different length-scales. *Appl. Phys. Lett.*, 99(23):231907.

[88] Mehta, D., Zhao, X., Bernal, E. A., and Wales, D. J. (2018). Loss surface of XOR artificial neural networks. *Phys. Rev. E*, 97(5):052307.

[89] Mei, S., Montanari, A., and Nguyen, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA*, 115:7665.

[90] Mezey, P. G. (1987). *Potential energy hypersurfaces*. Elsevier Science Ltd, Amsterdam.

[91] Middleton, T. F., Hernández-Rojas, J., Mortenson, P. N., and Wales, D. J. (2001). Crystals of binary Lennard-Jones solids. *Phys. Rev. B*, 64:184201.

[92] Miller, J. B., Zumbühl, D. M., Marcus, C. M., Lyanda-Geller, Y. B., Goldhaber-Gordon, D., Campman, K., and Gossard, A. C. (2003). Gate-Controlled Spin-Orbit Quantum Interference Effects in Lateral Transport. *Phys. Rev. Lett.*, 90(7):076807.

[93] Ming, Y., Lin, C.-T., Bartlett, S. D., and Zhang, W.-W. (2019). Quantum topology identification with deep neural networks and quantum walks. *npj Computational Materials*, 5(1):88.

[94] Mott, N. F. and Bohr, N. H. D. (1929). The scattering of fast electrons by atomic nuclei. *Proc. R. Soc. Lond. A*, 124(794):425.

[95] Munro, L. J. and Wales, D. J. (1999). Defect migration in crystalline silicon. *Phys. Rev. B*, 59:3969.

[96] Murrell, J. N. and Laidler, K. J. (1968). Symmetries of activated complexes. *Trans. Faraday. Soc.*, 64:371.

[97] Nagai, R., Akashi, R., and Sugino, O. (2020). Completing density functional theory by machine learning hidden messages from molecules. *npj Computational Materials*, 6:43.

[98] Nakajima, Y., Syers, P., Wang, X., Wang, R., and Paglione, J. (2015). One-dimensional edge state transport in a topological Kondo insulator. *Nature Phys.*, 12:213.

[99] Narayan, V., Nguyen, T.-A., Mansell, R., Ritchie, D. A., and Mussler, G. (2016). Interplay of spin–orbit coupling and superconducting correlations in germanium telluride thin films. *Phys. Status Solidi RRL*, 10(3):253.

[100] Narayan, V., Verpoort, P. C., Dann, J. R. A., Backes, D., Ford, C. J. B., Lanius, M., Jalil, A. R., Schüffelgen, P., Mussler, G., Conduit, G. J., and Grützmacher, D. (2019). Long-lived nonequilibrium superconductivity in a noncentrosymmetric Rashba semiconductor. *Phys. Rev. B*, 100:024504.

[101] Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J., and Seh, Z. W. (2020). Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence*, 2(3):161–170.

[102] Niblett, S. P., Biedermann, M., Wales, D. J., and de Souza, V. K. (2017). Pathways for diffusion in the potential energy landscape of the network glass former SiO_2. *J. Chem. Phys.*, 147(15):152726.

[103] Niblett, S. P., de Souza, V. K., Stevenson, J. D., and Wales, D. J. (2016). Dynamics of a molecular glass former: Energy landscapes for diffusion in ortho-terphenyl. *J. Chem. Phys.*, 145(2):024505.

[104] Nitta, J., Akazaki, T., Takayanagi, H., and Enoki, T. (1997). Gate Control of Spin-Orbit Interaction in an Inverted $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ Heterostructure. *Phys. Rev. Lett.*, 78:1335.

[105] Overhauser, A. W. (1953). Paramagnetic Relaxation in Metals. *Phys. Rev.*, 89:689.

[106] Pavlenko, N., Kopp, T., Tsymbal, E. Y., Sawatzky, G. A., and Mannhart, J. (2012). Magnetic and superconducting phases at the $LaAlO_3/SrTiO_3$ interface: The role of interfacial Ti 3*d* electrons. *Phys. Rev. B*, 85:020407.

[107] Pavlovskaia, M., Tu, K., and Zhu, S.-C. (2014). Mapping Energy Landscapes of Non-Convex Learning Problems. *arXiv e-prints*. arXiv:1410.0576.

[108] Pawley, G. S., Cochran, W., Cowley, R. A., and Dolling, G. (1966). Diatomic Ferroelectrics. *Phys. Rev. Lett.*, 17:753.

[109] Pham, H., Guan, M., Zoph, B., Le, Q., and Dean, J. (2018). Efficient Neural Architecture Search via Parameters Sharing. In Dy, J. and Krause, A., editors, *Proceedings of Machine Learning Research*, volume 80, pages 4095–4104. Machine Learning Research Press.

[110] Picozzi, S. (2014). Ferroelectric Rashba semiconductors as a novel class of multifunctional materials. *Frontiers in Phys.*, 2:10.

[111] Pincus, M. (1970). Letter to the Editor—A Monte Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems. *Operations Research*, 18(6):1225.

[112] Rabe, K. M. and Joannopoulos, J. D. (1987). Theory of the structural phase transition of GeTe. *Phys. Rev. B*, 36:6631.

[113] Rere, L. R., Fanany, M. I., and Arymurthy, A. M. (2015). Simulated Annealing Algorithm for Deep Learning. *Procedia Computer Sci.*, 72:137.

[114] Rigolin, G., Ortiz, G., and Ponce, V. H. (2008). Beyond the quantum adiabatic approximation: Adiabatic perturbation theory. *Phys. Rev. A*, 78(5):052508.

[115] Rinaldi, C., Varotto, S., Asa, M., Sławińska, J., Fujii, J., Vinai, G., Cecchi, S., Di Sante, D., Calarco, R., Vobornik, I., Panaccione, G., Picozzi, S., and Bertacco, R. (2018). Ferroelectric Control of the Spin Texture in GeTe. *Nano Lett.*, 18(5):2751.

[116] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv e-prints*. arXiv:1609.04747.

[117] Sagun, L., Bottou, L., and LeCun, Y. (2016). Eigenvalues of the Hessian in Deep Learning: Singularity and Beyond. *arXiv e-prints*. arXiv:1611.07476.

[118] Sarma, S. D., Freedman, M., and Nayak, C. (2015). Majorana zero modes and topological quantum computation. *npj Quantum Information*, 1:15001.

[119] Sarma, S. D. and Hwang, E. H. (2015). Screening and transport in 2D semiconductor systems at low temperatures. *Sci. Reports*, 5:16655.

[120] Sato, M. (2006). Nodal structure of superconductors with time-reversal invariance and $\mathbb{Z}_2$ topological number. *Phys. Rev. B*, 73:214502.

[121] Sato, M. and Fujimoto, S. (2009). Topological phases of noncentrosymmetric superconductors: Edge states, Majorana fermions, and non-Abelian statistics. *Phys. Rev. B*, 79:094504.

[122] Schonenberg, L. M., Verpoort, P. C., and Conduit, G. J. (2017). Effective-range dependence of two-dimensional Fermi gases. *Phys. Rev. A*, 96:023619.

[123] Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proc. Natl. Acad. Sci. USA*, 117(48):30033–30038.

[124] Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Math. Comput.*, 24:647.

[125] Shrestha, A. and Mahmood, A. (2019). Review of Deep Learning Algorithms and Architectures. *IEEE Access*, 7:53040.

[126] Sinova, J., Culcer, D., Niu, Q., Sinitsyn, N. A., Jungwirth, T., and MacDonald, A. H. (2004). Universal Intrinsic Spin Hall Effect. *Phys. Rev. Let.*, 92:126603.

[127] Sist, M., Kasai, H., Hedegaard, E. M. J., and Iversen, B. B. (2018). Role of vacancies in the high-temperature pseudodisplacive phase transition in GeTe. *Phys. Rev. B*, 97:094116.

[128] Sławińska, J., Di Sante, D., Varotto, S., Rinaldi, C., Bertacco, R., and Picozzi, S. (2019). Fe/GeTe(111) heterostructures as an avenue towards spintronics based on ferroelectric Rashba semiconductors. *Phys. Rev. B*, 99:075306.

[129] Smidman, M., Salamon, M. B., Yuan, H. Q., and Agterberg, D. F. (2017). Superconductivity and spin–orbit coupling in non-centrosymmetric materials: a review. *Reports on Prog. in Phys.*, 80(3):036501.

[130] Smith, S. L., Kindermans, P.-J., Ying, C., and Le, Q. V. (2018). Don't Decay the Learning Rate, Increase the Batch Size. In *6th International Conference on Learning Representations, ICLR 2018*.

[131] Smith, S. L. and Le, Q. V. (2017). A Bayesian Perspective on Generalization and Stochastic Gradient Descent. *arXiv e-prints*. arXiv:1710.06451.

[132] Stanev, V., Oses, C., Kusne, A. G., Rodriguez, E., Paglione, J., Curtarolo, S., and Takeuchi, I. (2018). Machine learning modeling of superconducting critical temperature. *npj Computational Materials*, 4:29.

[133] Stillinger, F. H. and Weber, T. A. (1984). Packing Structures and Transitions in Liquids and Solids. *Science*, 225(4666):983.

[134] Tanaka, Y., Yokoyama, T., Balatsky, A. V., and Nagaosa, N. (2009). Theory of topological spin current in noncentrosymmetric superconductors. *Phys. Rev. B*, 79:060505.

[135] Testelin, C., Eble, B., Bernardot, F., Karczewski, G., and Chamarro, M. (2008). Signature of the Overhauser field on the coherent spin dynamics of donor-bound electrons in a single CdTe quantum well. *Phys. Rev. B*, 77:235306.

[136] Tiwari, K. L., Coish, W. A., and Pereg-Barnea, T. (2017). Magnetoconductance signatures of chiral domain-wall bound states in magnetic topological insulators. *Phys. Rev. B*, 96:235120.

[137] Towsey, M., Alpsan, D., and Sztriha, L. (1995). Training a neural network with conjugate gradient methods. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 1, page 373. IEEE, New York.

[138] Trygubenko, S. A. and Wales, D. J. (2004a). A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.*, 120:2082.

[139] Trygubenko, S. A. and Wales, D. J. (2004b). Analysis of cooperativity and localization for atomic rearrangements. *J. Chem. Phys.*, 121:6689.

[140] Tsu, R., Howard, W., and Esaki, L. (1967). Optical properties of GeTe. *Solid State Commun.*, 5(3):167.

[141] Verpoort, P., MacDonald, P., and Conduit, G. (2018). Materials data validation and imputation with an artificial neural network. *Computational Materials Sci.*, 147:176.

[142] Verpoort, P. C., Lee, A. A., and Wales, D. J. (2020). Archetypal landscapes for deep neural networks. *Proc. Natl. Acad. Sci. USA*, 117(36):21857.

[143] Verpoort, P. C. and Narayan, V. (2020). Chirality relaxation in low-temperature strongly Rashba-coupled systems. *J. of Phys.: Condensed Matter*, 32(35):355704.

[144] Visintin, A. (1994). Chapter I. Genesis of Hysteresis. In John, F., Marsden, J. E., and Sirovich, L., editors, *Differential Models of Hysteresis*, Applied Mathematical Sciences, pages 12–31. Springer, Berlin.

[145] Wales, D. J. (2004). *Energy Landscapes: With Applications to Clusters, Biomolecules and Glasses*. Cambridge Molecular Science. Cambridge University Press, Cambridge.

[146] Wales, D. J. (2012). Decoding the energy landscape: extracting structure, dynamics and thermodynamics. *Phil. Trans. Roy. Soc. A*, 370:2877.

[147] Wales, D. J. (2018). Exploring Energy Landscapes. *Ann. Rev. Phys. Chem.*, 69:401.

[148] Wales, D. J. and Doye, J. P. K. (1997). Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A*, 101(28):5111.

[149] Wales, D. J. and Doye, J. P. K. (2003). Stationary points and dynamics in high-dimensional systems. *J. Chem. Phys.*, 119:12409.

[150] Wales, D. J., Miller, M. A., and Walsh, T. R. (1998). Archetypal energy landscapes. *Nature*, 394:758.

[151] Wang, T. E., Gu, Y., Mehta, D., Zhao, X., and Bernal, E. A. (2018). Towards Robust Deep Neural Networks. *arXiv e-prints*. arXiv:1810.11726.

[152] Wdowik, U. D., Parlinski, K., Rols, S., and Chatterji, T. (2014). Soft-phonon mediated structural phase transition in GeTe. *Phys. Rev. B*, 89:224306.

[153] Wolf, S. A., Awschalom, D. D., Buhrman, R. A., Daughton, J. M., von Molnár, S., Roukes, M. L., Chtchelkanova, A. Y., and Treger, D. M. (2001). Spintronics: A Spin-Based Electronics Vision for the Future. *Science*, 294(5546):1488.

[154] Wolgast, S., Eo, Y. S., Öztürk, T., Li, G., Xiang, Z., Tinsman, C., Asaba, T., Lawson, B., Yu, F., Allen, J. W., Sun, K., Li, L., Ç. Kurdak, Kim, D.-J., and Fisk, Z. (2015). Magnetotransport measurements of the surface states of samarium hexaboride using Corbino structures. *Phys. Rev. B*, 92:115110.

[155] Wu, L., Zhu, Z., and E, W. (2017). Towards Understanding Generalization of Deep Learning: Perspective of Loss Landscapes. *arXiv e-prints*. arXiv:1706.10239.

[156] Wu, M. W., Jiang, J. H., and Weng, M. Q. (2010). Spin dynamics in semiconductors. *Phys. Reports*, 493(2):61.

[157] Xu, M., Lei, Z., Yuan, J., Xue, K., Guo, Y., Wang, S., Miao, X., and Mazzarello, R. (2018). Structural disorder in the high-temperature cubic phase of GeTe. *Royal Soc. of Chem. Advances*, 8:17435.

[158] Yafet, Y. (1963). g Factors and Spin-Lattice Relaxation of Conduction Electrons. In Seitz, F. and Turnbull, D., editors, *Solid State Physics*, volume 14, pages 1–98. Academic Press, New York.

[159] Zeng, Y., Xiao, P., and Henkelman, G. (2014). Unification of algorithms for minimum mode optimization. *J. Chem. Phys.*, 140:044115.

[160] Zhai, H. (2015). Degenerate quantum gases with spin–orbit coupling: a review. *Reports on Prog. in Phys.*, 78(2):026001.

[161] Zhang, Y., Saxe, A. M., Advani, M. S., and Lee, A. A. (2018). Energy–entropy competition and the effectiveness of stochastic gradient descent in machine learning. *Molecular Phys.*, 116(21-22):3214.

[162] Zheng, L. and Das Sarma, S. (1996). Coulomb scattering lifetime of a two-dimensional electron gas. *Phys. Rev. B*, 53:9964.

[163] Žutić, I., Fabian, J., and Sarma, S. D. (2004). Spintronics: Fundamentals and applications. *Rev. Mod. Phys.*, 76:323.

Appendix A

A.1 First and second-order derivatives of the loss function

While the computation of the loss-function derivatives is conceptually simple, the practical implementation of the second-order derivatives is not standard in most ML codes, and its integration in the energy landscape exploration toolkit is an important contribution of this work, which is why a detailed description is provided here.

A.1.1 First-order derivatives

We start off by computing the first-order derivatives of the loss function, $\frac{\mathrm{d}L}{\mathrm{d}w_{ij}^{l}}$. We can write

$$\frac{\mathrm{d}}{\mathrm{d}w_{ij}^{l}} = \frac{\mathrm{d}a_i^l}{\mathrm{d}w_{ij}^{l}}\frac{\partial}{\partial a_i^l} = z_j^{l-1}\frac{\partial}{\partial a_i^l}\,. \tag{A.1}$$

Furthermore, we can define $\sigma_i^l = \frac{\partial L}{\partial a_i^l}$. For all $l \in \{1,\ldots,H+1\}$, these can be computed as

$$\sigma_i^l = \frac{\partial L}{\partial a_i^l} = \sum_j \frac{\partial L}{\partial a_j^{l+1}}\frac{\partial a_j^{l+1}}{\partial a_i^l} = \sum_j \sigma_j^{l+1}\frac{\partial a_j^{l+1}}{\partial z_i^l}\frac{\partial z_i^l}{\partial a_i^l} = \sum_j \sigma_j^{l+1} w_{ji}^{l+1} f'(a_i^l), \tag{A.2}$$

and so the σ_i^l can be calculated iteratively from the σ_i^{l+1}. We define $v_{ji}^l = w_{ji}^l f'(a_i^{l-1})$ so that we can write

$$\sigma_i^l = \sum_j \sigma_j^{l+1} v_{ji}^{l+1}\,. \tag{A.3}$$

This iteration can be computed if the σ_i^{H+1} are known. For the two loss functions defined in Sec. 2.1.2, it is

$$\sigma_i^{H+1} = 2\,(a_i^{H+1} - y_i) \quad \text{and} \quad \sigma_i^{H+1} = e^{a_i^{H+1}} \;/\; \sum_{a=0}^{3} e^{a_a^{H+1}} - \delta_{ic}\,. \tag{A.4}$$

Hence, the first-order derivatives are given by

$$\frac{\mathrm{d}L}{\mathrm{d}w_{ij}^{l}} = \sigma_i^l z_j^{l-1} \tag{A.5}$$

A.1.2 Second-order derivatives

A general methodology for calculating the second-order derivatives of a multi-layer perceptron was first reported by Bishop [18]. Here, we derive concrete expressions for the second-order derivatives of the DNN architecture defined in Sec. 2.1.1 that can easily be implemented in computational code.

We aim to calculate the second-order derivative of the loss function, $\frac{\mathrm{d}^2L}{\mathrm{d}w_{ij}^l\,\mathrm{d}w_{mn}^{l'}}$. Next for $1 \leq l' \leq l \leq H+1$, we compute

$$\frac{\mathrm{d}^2L}{\mathrm{d}w_{ij}^l\,\mathrm{d}w_{mn}^{l'}} \tag{A.6}$$

$$= \frac{\mathrm{d}}{\mathrm{d}w_{mn}^{l'}}\left(\frac{\mathrm{d}L}{\mathrm{d}w_{ij}^l}\right) \tag{A.7}$$

$$= z_n^{l'-1}\frac{\partial}{\partial a_m^{l'}}\left(z_j^{l-1}\sigma_i^l\right) \tag{A.8}$$

$$= z_j^{l-1}z_n^{l'-1}\frac{\mathrm{d}\sigma_i^l}{\mathrm{d}a_m^{l'}} + \sigma_i^l z_n^{l'-1}\frac{\mathrm{d}z_j^{l-1}}{\mathrm{d}a_m^{l'}} \tag{A.9}$$

$$= z_j^{l-1}z_n^{l'-1}\frac{\mathrm{d}\sigma_i^l}{\mathrm{d}a_m^{l'}} + \sigma_i^l z_n^{l'-1}f'(a_j^{l-1})\frac{\mathrm{d}a_j^{l-1}}{\mathrm{d}a_m^{l'}} \tag{A.10}$$

$$= z_j^{l-1}z_n^{l'-1}b_{im}^{ll'} + \sigma_i^l z_n^{l'-1}f'(a_j^{l-1})\,g_{jm}^{l-1,l'}\,, \tag{A.11}$$

where we have defined $b_{im}^{ll'} = \frac{\mathrm{d}\sigma_i^l}{\mathrm{d}a_m^{l'}}$ and $g_{jm}^{ll'} = \frac{\mathrm{d}a_j^l}{\mathrm{d}a_m^{l'}}$. For $l' > l$, it is $g_{jm}^{ll'} = 0$, for $l' = l$ it is $g_{jm}^{ll'} = \delta_{jm}$, and for $l' = l-1$ we find

$$g_{jm}^{ll'} = \frac{\mathrm{d}a_j^l}{\mathrm{d}a_m^{l-1}} = w_{jm}^l f'(a_m^{l-1}) = v_{jm}^l\,, \tag{A.12}$$

and we can use

$$g^{ll'}_{jm} = \frac{\mathrm{d}a^l_j}{\mathrm{d}a^{l'}_m} = \sum_k \frac{\mathrm{d}a^l_j}{\mathrm{d}a^{l-1}_k} \frac{\mathrm{d}a^{l-1}_k}{\mathrm{d}a^{l'}_m} = \sum_k f'(a^{l-1}_k)\, w^l_{jk}\, g^{l-1,l'}_{km} = \sum_k v^l_{jk}\, g^{l-1,l'}_{km} \tag{A.13}$$

to compute each entry of g iteratively. Next, we find that

$$b^{ll'}_{im} = \frac{\mathrm{d}\sigma^l_i}{\mathrm{d}a^{l'}_m} \tag{A.14}$$

$$= \frac{\mathrm{d}}{\mathrm{d}a^{l'}_m} \left(\sum_j \sigma^{l+1}_j w^{l+1}_{ji} f'(a^l_i) \right) \tag{A.15}$$

$$= \sum_j b^{l+1,l'}_{jm} w^{l+1}_{ji} f'(a^l_i) + \sum_j \sigma^{l+1}_j w^{l+1}_{ji} f''(a^l_i) g^{ll'}_{im}\,. \tag{A.16}$$

We define $u^l_{ji} = w^l_{ji} f''(a^{l-1}_i)$, so that we can write

$$b^{ll'}_{im} = \sum_j b^{l+1,l'}_{jm} v^{l+1}_{ji} + \sum_j \sigma^{l+1}_j u^{l+1}_{ji} g^{ll'}_{im}\,. \tag{A.17}$$

Because the definition of $b^{ll'}$ is symmetric with respect to interchanging $l \leftrightarrow l'$, we can also write,

$$b^{ll'}_{im} = \sum_j b^{l,l'+1}_{ij} v^{l'+1}_{jm} + \sum_j \sigma^{l'+1}_j u^{l'+1}_{ji} g^{l'l}_{mi}\,. \tag{A.18}$$

We will use this formula to compute all $b^{H+1,l'}$ from $b^{H+1,H+1}$. In that case, the second term vanishes because $g^{l'l}_{mi} = 0$ for $l' < l$. Hence,

$$b^{H+1,l'}_{im} = \sum_j b^{H+1,l'+1}_{ij} v^{l'+1}_{jm}\,. \tag{A.19}$$

Finally, it is

$$b^{H+1,H+1}_{im} = 2\,\delta_{im} \quad \text{and} \quad b^{H+1,H+1}_{im} = -\,e^{a^{H+1}_i}\, \frac{e^{a^{H+1}_m} - \delta_{im} \sum_{a=0}^{3} e^{a^{H+1}_a}}{\left(\sum_{a=0}^{3} e^{a^{H+1}_a}\right)^2}\,. \tag{A.20}$$

A.2 Creation of the LJAT19 dataset

In this appendix, we provide details of the creation of the LJAT19 dataset and the visualisations of the corresponding classification problem, which we make use of in Chap. 3.

A.2.1 Predicting the outcome of geometry optimisation for an atomic cluster

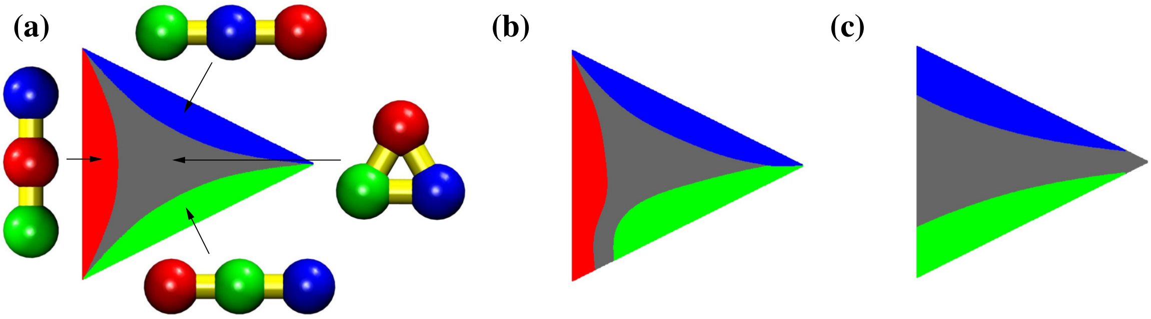
The dataset employed in this work is based on a geometry optimisation problem and has been used for benchmarking problems considered in several previous contributions that employed single hidden-layer neural networks [11, 12, 37, 30]. The geometry optimisation problem is defined by a triatomic system with pairwise Lennard-Jones [66] and three-body Axilrod-Teller [7] interactions. Its locally optimal geometries are three permutational isomers of a linear minimum with all three atoms in a line distinguished by the central atom and one additional geometry for an equilateral triangle with D_{3h} symmetry. The total potential energy for this LJAT$_3$ cluster is given by

$$V = 4\varepsilon \sum_{i<j} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] + Z \sum_{i<j<k} \left[\frac{1 + 3\cos\theta_1 \cos\theta_2 \cos\theta_3}{(r_{ij} r_{ik} r_{jk})^3} \right], \qquad \text{(A.21)}$$

where θ_1, θ_2, and θ_3 are the internal angles of the triangle formed by atoms i, j, and k. The distance between atoms i and j is r_{ij}, and Z is a parameter that weights the contribution of the three-body term. For $Z = 2$ the linear minima have potential energy $V = -2.219\,\varepsilon$, and the triangle lies slightly higher with $V = -2.185\,\varepsilon$. For the triangle $r_{12} = r_{13} = r_{23} = 1.16875\,\sigma$, and in the linear minima the nearest-neighbour distances are both $1.10876\,\sigma$.

A four-fold classification problem can be defined based on the input coordinates of this geometry optimisation problem and its four possible outcomes. The class index, which ranges from 1 to 4, can be mapped out across all positive values of the three input variables r_{12}, r_{13}, and r_{23}, and the supports of the class indices equal the catchment basins of the respective local minima of the LJAT$_3$ cluster. In order to turn this classification problem into a learnable dataset, we generate random starting geometries that distribute the three atoms in a cube of side length L, and we evaluate the true corresponding class index from geometry optimisation by determining which of the four structures the initial state relaxes to. The LJAT19 dataset created for this work is based on 200 000 geometry minimisations for starting positions in a cube with $L = 1.385\,\sigma$. Half of each dataset is used for training and the other half for testing, where appropriate. A similar dataset with 10 000 minimisations and $L = 2\sqrt{3}\,\sigma$ was employed in previous work [11, 12, 37, 30].

When training the DNN architectures in Chap. 3, we restrict the inputs to just two of the three interparticle distances, i.e. providing r_{12} and r_{13} only while omitting r_{23}. This makes the prediction problem harder, as it can no longer be known with certainty what the correct outcome is. The best outcome to expect from a converged DNN architecture is a prediction of the classification index with the highest marginal probability.



Reprinted with permission from Ref. [142], P.C. Verpoort et al., Proc. National Academy of Sci. USA 117, 21857 (2020). © *2020 National Academy of Sciences.*

Fig. A.1 Graphical representation of the LJAT$_3$ classification problem. **(a)** Coloured according to the true outcome determined by geometry optimisation for the LJAT$_3$ cluster. The four optimal atomic configurations are associated with their corresponding basins of attraction. **(b)** Coloured according to the predictions for the global minimum of a single hidden-layer neural network with 3 hidden nodes and 100 training data confined in the (x,y) plane (AUC 0.98 from the corresponding test set). **(c)** Coloured according to the predictions for the global minimum of a single hidden-layer neural network with 10 hidden nodes trained on 100 000 training data from the LJAT19 dataset (AUC 0.79 from corresponding test set).

We find the benchmark presented in this appendix appealing because we can generate arbitrary amounts of training and testing data, because the classification problem has a clear physical interpretation, and because of the practical importance of the configuration volumes themselves. The training and testing datasets with 100 000 entries each, which are generated from this benchmark, are therefore used heavily in Chap. 3, yet we also validate our key results with real-life data examples in Sec. 3.2.7.

A.2.2 Visualisation of solutions

We aim to visualise prediction functions of DNN models trained with the LJAT19 dataset. This is accomplished by plotting the classification index in a representative subspace of the full three-dimensional $[r_{12}, r_{13}, r_{23}]$ space. Specifically, we project coordinates onto the plane $r_{12} + r_{13} + r_{23} = 3r_e$ with $r_e = 2^{1/6}$ being the equilibrium bond length in a dimer and in the equilateral triangle minimum. This plane is perpendicular to the $\{1,1,1\}$ direction and spanned by the unit vectors $\hat{\mathbf{v}}_1 = (1,1,-2)/\sqrt{6}$ and $\hat{\mathbf{v}}_2 = (1,-1,0)/\sqrt{2}$. The projected coordinates can be defined as $x = (r_{12} + r_{13} - 2r_{23})/\sqrt{6}$ and $y = (r_{12} - r_{13})/\sqrt{2}$, such that each point in the $[r_{12}, r_{13}, r_{23}]$ space can be assigned an (x,y) coordinate.

Fig. A.1 shows different visualisations of the LJAT19 class index on the projected plane. Fig. A.1(a) colours each point on the (x,y) grid according to the true class index of the corresponding point on the projected plane. The equilateral triangle is coloured in grey, while the three linear minima with atoms 1, 2, and 3 in the centre are coloured red, green, and blue, respectively. Fig. A.1(b) presents the results of a single hidden-layer neural network trained data confined to the projected plane. In this confined space, the knowledge of r_{12} and r_{23} is sufficient to determine the correct class index, which is why the visualised solution comes close to the correct one in Fig. A.1(a). Finally, Fig. A.1(c) shows the prediction of a neural network trained on the LJAT19 dataset, where points in the training dataset are distributed all across the $[r_{12}, r_{13}, r_{23}]$ space. Consequently, the network cannot distinguish between the red and the grey solution. The presented visualisation in Fig. A.1(c) is hence the best outcome we can expected neural networks to converge to that have been trained on data from the LJAT19 dataset with only two inputs provided.

While this graphical visualisation only represents a small subspace of all possible input configurations from the three-dimensional $[r_{12}, r_{13}, r_{23}]$ space, it can serve as a useful tool to compare DNN models and their predictions and provide valuable insight into how well a particular DNN model performs. This visualisation can be plotted for any DNN configuration, including minima, TSs, and generally any point in the LFL. Its main application in this work is in Fig. 3.2, where it is employed to visualise the performance of the global minimum of different DNN architectures for varying number of training data, N_{data}. Interestingly, the complexity of the patterns in these visualisations suggest a higher expressibility of those DNNs with higher H, as further outlined in the main text.

A.3 Identification of permutational DNN isomers from loss value difference

As discussed in Sec. 3.1.2, our analysis of the LFL in DNNs relies on the identification of degeneracies (or isomers) arising from permutational degeneracies when searching for minima and TSs. Because of the difficulties associated with computing the distance in weight space between two points while considering all permutational symmetries, the present work takes a simplified approach to identify these: minima are treated as identical if their difference in training loss value is below a certain threshold, namely the loss-difference tolerance, ΔL_{tol}. This assumption is necessary when no other methods of identifying permutational isomers is available because we would otherwise overestimate the number of minima (potentially in quite a significant way, given the factorial growth of the DNN symmetries with the number of nodes in a hidden layer, as explained at the end of Sec. 2.1.1). The simplified approach we employ however also holds the risk of miss-identifying minima, hence discarding newly discovered minima as permutational isomers of known minima when the new ones are in fact separate minima.

Choosing a sensible value of ΔL_{tol} is not straightforward. This number has to be somewhat greater than the loss convergence tolerance of all minima. Given the finite convergence tolerance of the training loss of minima due to various computational constraints, ΔL_{tol} should be large enough so that two representations of the same minimum are not identified as independent minima because their loss values were not converged sufficiently. Choosing ΔL_{tol} as too large will generally result in accidental identification of minima. This is particularly likely to happen when minima of the LFL occur at high density such that the difference ΔL in their training loss values is less than ΔL_{tol}. Following some initial estimation, the number ΔL_{tol} was set to be either 10^{-6} or 10^{-7}. This section continues to study whether that choice of ΔL_{tol} is in fact sufficient given the density of minima.

In order to analyse the extent of potential deficiencies arising from too low ΔL_{tol}, we plot the difference ΔL of neighbouring minima discovered from the procedure described in Sec. 3.1.2 as a function of the training loss of these minima, L_{train}. We discover that in those examples with few minima, i.e. any of the cases with $H = 1$ as well as those with $N_{\text{data}} \geq 2000$ for $H = 2$ and $N_{\text{data}} \geq 10000$ for $H = 3$, the distance between discovered minima in training loss space is greater than the loss-difference tolerance, ΔL_{tol}, hence it is unlikely that we have missed minima in those cases. On the contrary, cases with $N_{\text{data}} \leq 1000$ for $H = 2$ and $N_{\text{data}} \leq 2000$ for $H = 3$ have minima with ΔL very close to ΔL_{tol}. In fact the spectrum shows a clear cut-off at ΔL_{tol}. This indicates the miss-identification of dissimilar minima as permutational isomers due to an insufficiently large choice of ΔL_{tol}.

This analysis indicates that the results obtained for the latter cases of the LFLs using the LJAT19 dataset are incomplete: some minima have gone undiscovered because they were falsely identified as permutational isomers of already existent minima in the database. Future work has to either incorporate a sufficiently small ΔL_{tol} or rely on additional, more advanced procedures for checking permutational isomers, such as through normal ordering of nodes (as alluded to in Sec. 3.1.2) in order to establish a complete picture of the LFL in those high-density cases.

While not performed explicitly here, the analysis presented above does of course carry over to TSs as well, for which we expect to find similar results and consequences.

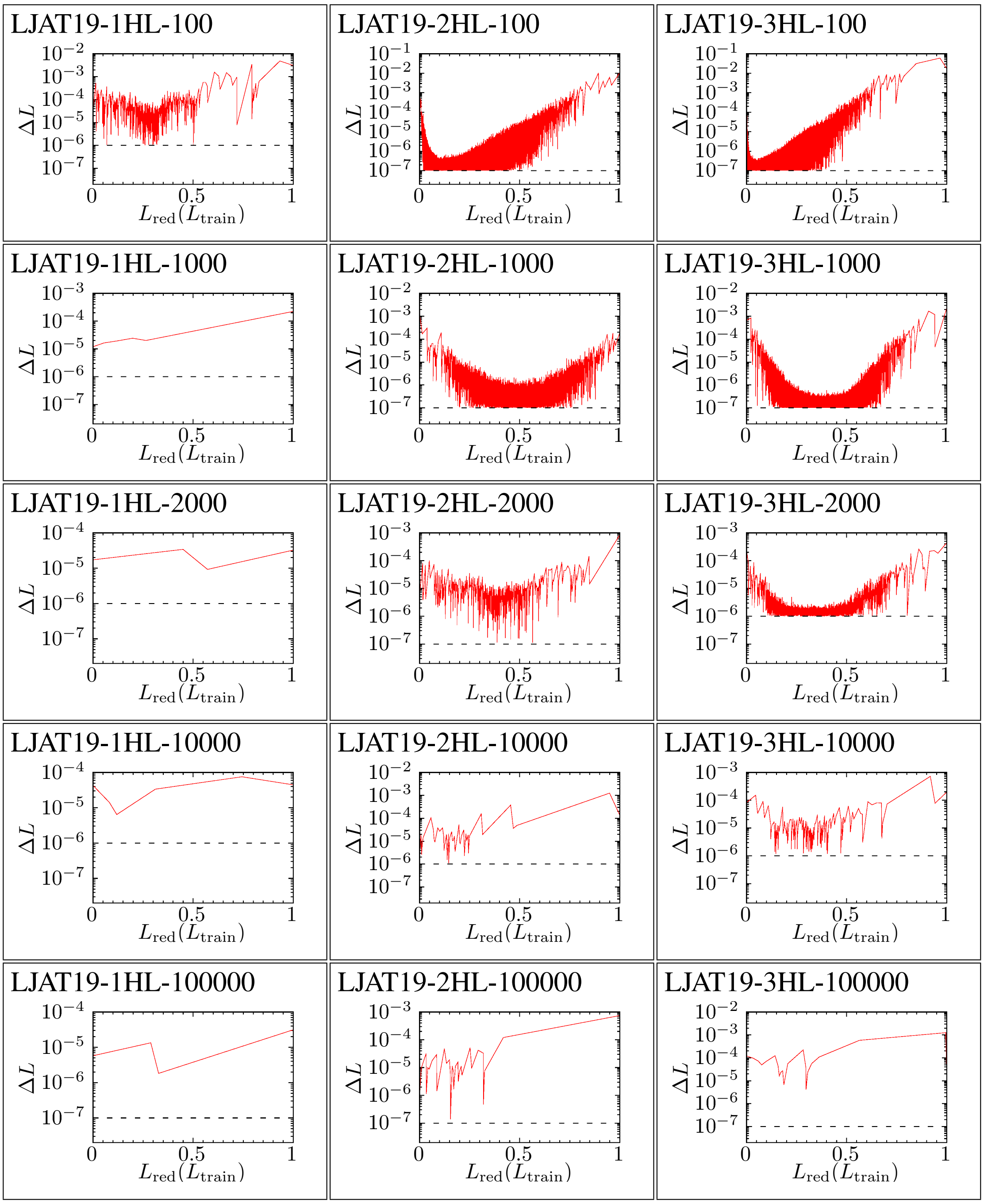


Fig. A.2 Training loss difference, ΔL, between neighbouring LFL minima as a function of the reduced training loss, $L_{\rm red}(L_{\rm train})$, for the LJAT19 dataset. The reduced loss is defined in the same way as in Fig. 3.4 and is computed as $L_{\rm red}(L) = \frac{L-L_{\rm min}}{L_{\rm max}-L_{\rm min}}$, where $L_{\rm max}$ is the maximal and $L_{\rm min}$ is the minimal loss value in the corresponding database of minima. The dashed line at the bottom of each graph indicates the loss-difference tolerance, $\Delta L_{\rm tol}$. Columns left to right: H taking values 1, 2, and 3. Rows top to bottom: $N_{\rm data}$ taking values 100, 1000, 2000, 10 000, and 100 000.

Appendix B

B.1 Differential equation for dynamical Fermi energies

This appendix section derives a coupled ordinary differential equation for the Fermi energies of the two Rashba bands that describes their dynamics under a constantly ramped external magnetic field.

We start from the energy dispersions of the Rashba bands, $\varepsilon_k^\pm$, subject to a magnetic field, B,

$$\varepsilon_k^\pm = \varepsilon_k^{\mathrm{f}} \pm \sqrt{b^2 + 4E_{\mathrm{R}}\varepsilon_k^{\mathrm{f}}}\,, \tag{B.1}$$

which is derived from Eq. (5.2), and where we have defined the free energy dispersion, $\varepsilon_k^{\mathrm{f}} = \hbar^2 k^2/2m$, and the magnetic field in the dimensions of energy, $b = g\mu_{\mathrm{B}}B$. This allows us to compute the value of the Fermi momentum, $k_{\mathrm{F}}^\pm$, for any given Fermi energy ε_{F}. Instead of solving for the Fermi momentum, we instead solve for the value of the free energy dispersion at the Fermi momentum, $\varepsilon^{\mathrm{f}}_{k_{\mathrm{F}}^\pm}$, which yields

$$\varepsilon^{\mathrm{f}}_{k_{\mathrm{F}}^\pm} = \varepsilon_{\mathrm{F}}^\pm + 2E_{\mathrm{R}} \mp \sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^\pm}\,. \tag{B.2}$$

The total number of carriers $n^\pm$ is given by

$$n^\pm = \int \frac{\mathrm{d}^2k}{(2\pi)^2}\,\Theta(\varepsilon_{\mathrm{F}} - \varepsilon_k^\pm)\,, \tag{B.3}$$

where $\Theta(x)$ is the Heaviside function. By substituting $\varepsilon = \varepsilon_k^{\mathrm{f}}$, we find

$$n^\pm = \frac{m}{2\pi\hbar^2}\int_0^\infty \mathrm{d}\varepsilon\;\Theta(\varepsilon^{\mathrm{f}}_{k_{\mathrm{F}}^\pm} - \varepsilon) = \frac{m}{2\pi\hbar^2}\varepsilon^{\mathrm{f}}_{k_{\mathrm{F}}^\pm}\,. \tag{B.4}$$

We define the reduced carrier density $\nu^{\pm} = \frac{2\pi\hbar^2}{m} n^{\pm}$, which has the dimension of energy, such that

$$\nu^{\pm} = \varepsilon_{\mathrm{F}}^{\pm} + 2E_{\mathrm{R}} \mp \sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}\,. \tag{B.5}$$

This equation can be inverted to yield the Fermi energies $\varepsilon_{\mathrm{F}}^{\pm}$ from the carrier density $\nu^{\pm}$,

$$\varepsilon_{\mathrm{F}}^{\pm} = \nu^{\pm} \pm \sqrt{b^2 + 4E_{\mathrm{R}}\nu^{\pm}}\,. \tag{B.6}$$

As explained in Sec. 5.3.2, we define the total density, $\nu_{\mathrm{tot}} = \nu^+ + \nu^-$, and the chirality density, $\nu_{\mathrm{c}} = \nu^- - \nu^+$, as the time dependence of the occupation numbers $\nu^{\pm}$ are determined by a relaxation time approximation,

$$\frac{\partial \nu_{\mathrm{c}}}{\partial t} = -\frac{\nu_{\mathrm{c}} - \nu_{\mathrm{c}}^{\mathrm{eq.}}}{\tau}\,. \tag{B.7}$$

Using $\nu^{\pm} = \frac{\nu_{\mathrm{tot}} \mp \nu_{\mathrm{c}}}{2}$, we can write

$$\frac{\partial \nu^{\pm}}{\partial t} = \pm\frac{1}{2}\frac{\nu_{\mathrm{c}} - \nu_{\mathrm{c}}^{\mathrm{eq.}}}{\tau}\,. \tag{B.8}$$

It is

$$\nu_{\mathrm{c}} = \nu^- - \nu^+ = \varepsilon_{\mathrm{F}}^- - \varepsilon_{\mathrm{F}}^+ + \sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^-} + \sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^+}\,, \tag{B.9}$$

and

$$\nu_{\mathrm{c}}^{\mathrm{eq.}} = \nu_{\mathrm{c}}\Big|_{\varepsilon_{\mathrm{F}}^{\pm} = \varepsilon_{\mathrm{F}}^{\mathrm{eq.}}} = 2\sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}}\,. \tag{B.10}$$

We now turn to the expression defined in Eq. (5.3) that governs the dynamics of the Fermi energy. This can now be written as

$$\frac{\mathrm{d}\varepsilon_{\mathrm{F}}^{\pm}}{\mathrm{d}t} = \frac{\partial \varepsilon_{\mathrm{F}}^{\pm}}{\partial b}\dot{b} + \frac{\partial \varepsilon_{\mathrm{F}}^{\pm}}{\partial \nu^{\pm}}\frac{\partial \nu^{\pm}}{\partial t}\,, \tag{B.11}$$

where $\dot{b}$ is a constant. Using

$$\frac{\partial \varepsilon_{\mathrm{F}}^{\pm}}{\partial b} = \pm\frac{b}{\sqrt{b^2 + 4E_{\mathrm{R}}\nu^{\pm}}} = \frac{b}{\varepsilon_{\mathrm{F}}^{\pm} - \nu^{\pm}} = \frac{b}{-2E_{\mathrm{R}} \pm \sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}} \tag{B.12}$$

and

$$\frac{\partial \varepsilon_{\mathrm{F}}^{\pm}}{\partial \nu^{\pm}} = 1 + \frac{2E_{\mathrm{R}}}{b}\frac{\partial \varepsilon_{\mathrm{F}}^{\pm}}{\partial b} = 1 + \frac{2E_{\mathrm{R}}}{-2E_{\mathrm{R}} \pm \sqrt{b^2 + 4E_{\mathrm{R}}^2 + 4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}}\,, \tag{B.13}$$

we can write

$$\begin{aligned}\frac{\mathrm{d}\varepsilon_{\mathrm{F}}^{\pm}}{\mathrm{d}t} =& \frac{b\dot{b}}{-2E_{\mathrm{R}} \pm \sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}} \\ & \pm\frac{1}{2\tau}\left(1+\frac{2E_{\mathrm{R}}}{-2E_{\mathrm{R}} \pm \sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\pm}}}\right) \times \left(\varepsilon_{\mathrm{F}}^{-} - \varepsilon_{\mathrm{F}}^{+} + 2\sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}}\right. \\ & \left. - \sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{+}} - \sqrt{b^2+4E_{\mathrm{R}}^2+4E_{\mathrm{R}}\varepsilon_{\mathrm{F}}^{-}}\right),\end{aligned} \tag{B.14}$$

which is the equation reported in Sec. 5.3.2.

B.2 Condition on the excess occupation

This appendix section aims to derive an expression for the upper bound of the excess occupation δ from the Rashba band dispersion ε_k and the value of the equilibrium Fermi energy $\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}$.

The Fermi energies $\varepsilon_{\mathrm{F}}^{\pm}$ from the carrier density $\nu^{\pm}$ can be obtained from Eq. (B.6) by setting $B=0$, which yields

$$\varepsilon_{\mathrm{F}}^{\pm} = \nu^{\pm} \pm 2\sqrt{E_{\mathrm{R}}\nu^{\pm}}. \tag{B.15}$$

Therefore, given occupations $\nu^{\pm} = \nu^{\pm,\mathrm{eq.}}(1\pm\delta)$, we can calculate the Fermi energy difference to be

$$\Delta\varepsilon_{\mathrm{F}} = \varepsilon_{\mathrm{F}}^{+} - \varepsilon_{\mathrm{F}}^{-} \tag{B.16}$$

$$\begin{aligned}&= \nu^{+,\mathrm{eq.}}(1+\delta) - \nu^{-,\mathrm{eq.}}(1-\delta) \\ &\quad + 2\sqrt{E_{\mathrm{R}}}\left(\sqrt{\nu^{+,\mathrm{eq.}}(1+\delta)} + \sqrt{\nu^{-,\mathrm{eq.}}(1-\delta)}\right)\end{aligned} \tag{B.17}$$

$$\begin{aligned}&= 2\delta(\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}+2E_{\mathrm{R}}) - 4\sqrt{E_{\mathrm{R}}^2+\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}E_{\mathrm{R}}} + 2\sqrt{E_{\mathrm{R}}} \\ &\quad \times\left[\sqrt{1+\delta}\sqrt{\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}+2E_{\mathrm{R}}-2\sqrt{E_{\mathrm{R}}^2+\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}E_{\mathrm{R}}}}\right. \\ &\quad \left. + \sqrt{1-\delta}\sqrt{\varepsilon_{\mathrm{F}}+2E_{\mathrm{R}}+2\sqrt{E_{\mathrm{R}}^2+\varepsilon_{\mathrm{F}}E_{\mathrm{R}}}}\right],\end{aligned} \tag{B.18}$$

and, by expanding $\sqrt{1+\delta}+\sqrt{1-\delta}$ as a series around $\delta=0$, we find

$$\begin{aligned}
&= 2\delta(\varepsilon_{\mathrm{F}}+E_{\mathrm{R}})+4\sqrt{E_{\mathrm{R}}} \\
&\times\left[\sqrt{E_{\mathrm{R}}+\varepsilon_{\mathrm{F}}^{\mathrm{eq.}}}\sum_{k=1}^{\infty}\binom{1/2}{2k}\delta^{2k}\right. \\
&\left.-\sqrt{E_{\mathrm{R}}}\quad\sum_{k=1}^{\infty}\binom{1/2}{2k+1}\delta^{2k+1}\right].
\end{aligned} \tag{B.19}$$

When δ is small, the series expansion can be truncated after the first element, which yields the result reported in Sec. 6.2.

B.3 Computation of relaxation-time constants

B.3.1 General considerations

In this appendix, we provide a detailed explanation of our calculations of the Boltzmann-transport scattering integrals that are discussed in Sec. 6.3 of the main text. We start by providing some general calculations that hold for both carrier-impurity and carrier-carrier scattering, and we continue by calculating expressions for the relaxation-time constants for the two cases explicitly.

Boltzmann equation and scattering integrals

We consider the Boltzmann equation,

$$\frac{\partial f_{\mathbf{k}_1}^{\pm}}{\partial t}=I_{\mathrm{ci}}[f_{\mathbf{k}_1}^{\pm}]+I_{\mathrm{cc}}[f_{\mathbf{k}_1}^{\pm}]\,, \tag{B.20}$$

where the indices ci and cc refer to the carrier-impurity and carrier-carrier contributions of the scattering integral, respectively. These are given by

$$\begin{aligned}
I_{\mathrm{ci}}=\sum_{\mathbf{k}_2}\ &\big(w^{\mathrm{car-imp}}_{(\mathbf{k}_2\mp)\to(\mathbf{k}_1\pm)}\ f_{\mathbf{k}_2}^{\mp}[1-f_{\mathbf{k}_1}^{\pm}] \\
&-w^{\mathrm{car-imp}}_{(\mathbf{k}_1\pm)\to(\mathbf{k}_2\mp)}\ f_{\mathbf{k}_1}^{\pm}[1-f_{\mathbf{k}_2}^{\mp}]\big)
\end{aligned} \tag{B.21}$$

and

$$I_{\mathrm{cc}} = \sum_{\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} \big(w^{\mathrm{car-car}}_{(\mathbf{k}_3\mathbf{k}_4\mp)\to(\mathbf{k}_1\mathbf{k}_2\pm)}\, f^{\mp}_{\mathbf{k}_3} f^{\mp}_{\mathbf{k}_4} [1-f^{\pm}_{\mathbf{k}_1}][1-f^{\pm}_{\mathbf{k}_2}] \\ - w^{\mathrm{car-car}}_{(\mathbf{k}_1\mathbf{k}_2\pm)\to(\mathbf{k}_3\mathbf{k}_4\mp)}\, f^{\pm}_{\mathbf{k}_1} f^{\pm}_{\mathbf{k}_2} [1-f^{\mp}_{\mathbf{k}_3}][1-f^{\mp}_{\mathbf{k}_4}] \big)\,. \tag{B.22}$$

We have neglected all terms that conserve the Rashba-band index of each particle or induce an exchange of particles between the bands (as the latter will not lead to a decay of the carrier imbalance between the two Rashba bands). We will make up for this by assuming that *intra*-band scattering events are so quick that they will relax each individual band into *local* thermal equilibrium on a timescale that is immediate compared to the *inter*-band processes we study. We note that there exist additional inter-carrier scattering processes that conserve the Rashba-band index of one carrier but not of the other. These however are (due to considerations based on momentum conservation) absent when $k_{\mathrm{R}} > k_{\mathrm{F}}^{+}$ and remain weak as long as $k_{\mathrm{R}} \approx k_{\mathrm{F}}^{+}$, which we assume to be the case and which is the case for Rashba materials where the Rashba energy scale is comparable to the Fermi energy.

Defining the chirality density

We define the total particle density, $n = n^{+} + n^{-}$, and the total chirality density, $C = n^{-} - n^{+}$ (which is defined to be a positive number, as there are more particles in the lower Rashba band). Using the distribution function $f^{\pm}_{\mathbf{k}}$, we can write the densities as $n^{\pm} = \frac{1}{\mathscr{V}} \sum_{\mathbf{k}_1} f^{\pm}_{\mathbf{k}_1}$, and, therefore, we can express C as

$$C = \frac{1}{\mathscr{V}} \sum_{\mathbf{k}_1} \left(f^{-}_{\mathbf{k}_1} - f^{+}_{\mathbf{k}_1} \right) \tag{B.23}$$

and its derivative with respect to time t as

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \frac{1}{\mathscr{V}} \sum_{\mathbf{k}_1} \left(\frac{\partial f^{-}_{\mathbf{k}_1}}{\partial t} - \frac{\partial f^{+}_{\mathbf{k}_1}}{\partial t} \right), \tag{B.24}$$

where $\mathscr{V}$ is the area of the system.

We insert the carrier-impurity contribution of the scattering integral in Eq. (B.21) and the carrier-carrier contribution in Eq. (B.22) into Eq. (B.24). By, in the second term, renaming $\mathbf{k}_1 \mapsto \mathbf{k}_2$ and $\mathbf{k}_2 \mapsto \mathbf{k}_1$ for the carrier-impurity case and $\mathbf{k}_1, \mathbf{k}_2 \mapsto \mathbf{k}_3, \mathbf{k}_4$ and $\mathbf{k}_3, \mathbf{k}_4 \mapsto \mathbf{k}_1, \mathbf{k}_2$ in the carrier-carrier case and then making use of the reversibility of the microscopic processes, i.e. $w^{\text{car}-\text{imp}}_{(\mathbf{k}_2+)\to(\mathbf{k}_1-)} = w^{\text{car}-\text{imp}}_{(\mathbf{k}_1-)\to(\mathbf{k}_2+)}$ and $w^{\text{car}-\text{car}}_{(\mathbf{k}_3\mathbf{k}_4+)\to(\mathbf{k}_1\mathbf{k}_2-)} = w^{\text{car}-\text{car}}_{(\mathbf{k}_1\mathbf{k}_2-)\to(\mathbf{k}_3\mathbf{k}_4+)}$, we find

$$\left.\frac{\mathrm{d}C}{\mathrm{d}t}\right|_{\text{ci}} = \frac{2}{\mathscr{V}} \sum_{\mathbf{k}_1\mathbf{k}_2} w^{\text{car}-\text{imp}}_{(\mathbf{k}_1-)\to(\mathbf{k}_2+)} (f^+_{\mathbf{k}_2} - f^-_{\mathbf{k}_1}) \tag{B.25}$$

and

$$\begin{aligned}\left.\frac{\mathrm{d}C}{\mathrm{d}t}\right|_{\text{cc}} = \frac{2}{\mathscr{V}} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} w^{\text{car}-\text{car}}_{(\mathbf{k}_1\mathbf{k}_2-)\to(\mathbf{k}_3\mathbf{k}_4+)} \Big(& f^+_{\mathbf{k}_3} f^+_{\mathbf{k}_4} [1 - f^-_{\mathbf{k}_1}][1 - f^-_{\mathbf{k}_2}] \\ - \; & f^-_{\mathbf{k}_1} f^-_{\mathbf{k}_2} [1 - f^+_{\mathbf{k}_3}][1 - f^+_{\mathbf{k}_4}] \Big) .\end{aligned} \tag{B.26}$$

Inducing a Fermi-level detuning

We use the Rashba energy dispersion $\varepsilon^{\pm}_{k} = \frac{\hbar^2(k \pm k_{\text{R}})^2}{2m} - E_{\text{R}}$ (where $E_{\text{R}} = \hbar^2 k_{\text{R}}^2/2m$ is the Rashba energy) as shown in Fig. 6.1, although we shift all energies by E_{R} so that we can neglect the offset and assume the entire energy dispersion to take positive values.

We assume the equilibrium Fermi momentum $k_{\text{F}}^{\text{eq.}\pm} = k_{\text{F}}^{0} \mp k_{\text{R}}$ and the equilibrium chemical potential $\mu^{\text{eq.}} = \hbar^2(k_{\text{F}}^0)^2/2m$. A chirality non-equilibrium distribution that conserves the total density $n = n^+ + n^-$ is induced by letting $\left(k_{\text{F}}^{\pm}\right)^2 = \left(k_{\text{F}}^{\text{eq.}\pm}\right)^2 \mp \left(\delta k\right)^2$. By defining $\delta\mu = \hbar^2(\delta k)^2/2m$, we can express the non-equilibrium chemical potentials as

$$\mu^{\pm} = \mu^{\text{eq.}} \mp \frac{1}{1 \mp \sqrt{\frac{E_{\text{R}}}{\mu^{\text{eq.}}}}} \delta\mu + O(\delta\mu^2) . \tag{B.27}$$

We replace the distribution functions $f^{\pm}_{\mathbf{k}}$ in Eqs. (B.23), (B.25), and (B.26) by Fermi distributions $f_{\mu^{\pm}}(\varepsilon^{\pm}_{\mathbf{k}})$ with chemical potential $\mu^{\pm}$, which can, by using Eq. (B.27), be expanded to first order in $\delta\mu$ as

$$f_{\mu^{\pm}}(\varepsilon) = f(\varepsilon) \mp \chi^{\pm} \delta\mu \frac{\partial f(\varepsilon)}{\partial \mu} \tag{B.28}$$

and

$$[1 - f_{\mu^{\pm}}(\varepsilon)] = [1 - f(\varepsilon)] \mp \chi^{\pm} \delta\mu \frac{\partial [1 - f(\varepsilon)]}{\partial \mu} , \tag{B.29}$$

where we have defined $\chi^{\pm} = (1 \mp \sqrt{E_{\text{R}}/\mu^{\text{eq.}}})^{-1}$, and f is the equilibrium Fermi distribution. By using $\frac{\partial f}{\partial \mu} = -\frac{\partial f}{\partial \varepsilon} = \beta\, f(\varepsilon)\, [1 - f(\varepsilon)]$ and $\frac{\partial [1 - f(\varepsilon)]}{\partial \mu} = -\beta\, f(\varepsilon)\, [1 - f(\varepsilon)]$, we can rewrite

C as

$$C - C_{\text{eq.}} = + \frac{\delta\mu\,\beta}{\mathscr{V}} \sum_{\mathbf{k}_1} \Phi^0(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^+_{\mathbf{k}_1}) \tag{B.30}$$

and ${}^{\mathrm{d}C}/_{\mathrm{d}t}$ as

$$\left.\frac{\mathrm{d}C}{\mathrm{d}t}\right|_{\text{ci}} = -\frac{2\,\delta\mu\,\beta}{\mathscr{V}} \sum_{\mathbf{k}_1\mathbf{k}_2} \Phi^0(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^+_{\mathbf{k}_2}) \qquad \times \; w^{\text{car-imp}}_{(\mathbf{k}_1-)\to(\mathbf{k}_2+)}\,, \tag{B.31}$$

$$\left.\frac{\mathrm{d}C}{\mathrm{d}t}\right|_{\text{cc}} = -\frac{2\,\delta\mu\,\beta}{\mathscr{V}} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} \Phi^1(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^-_{\mathbf{k}_2}, \varepsilon^+_{\mathbf{k}_3}, \varepsilon^+_{\mathbf{k}_4}) \;\times\; w^{\text{car-car}}_{(\mathbf{k}_1\mathbf{k}_2-)\to(\mathbf{k}_3\mathbf{k}_4+)}\,. \tag{B.32}$$

The zeroth order term in $\delta\mu$ for $\frac{\mathrm{d}C}{\mathrm{d}t}$ vanishes as the equilibrium carrier configuration does not induce any changes of C. Furthermore, we have introduced $\Phi^0(\varepsilon_1, \varepsilon_2)$ and $\Phi^1(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$, which account for all distribution functions and which we define as

$$\Phi^0 = \chi^- f(\varepsilon_1)[1-f(\varepsilon_1)] + \chi^+ f(\varepsilon_2)[1-f(\varepsilon_2)]\,, \tag{B.33}$$

$$\begin{aligned} \Phi^1 = {} & f(\varepsilon_3)f(\varepsilon_4)[1-f(\varepsilon_1)][1-f(\varepsilon_2)] \\ & \times \Big(\chi^+\left([1-f(\varepsilon_3)]+[1-f(\varepsilon_4)]\right) \\ & \quad + \chi^-\left(f(\varepsilon_1)+f(\varepsilon_2)\right)\Big) \\ + {} & f(\varepsilon_1)f(\varepsilon_2)[1-f(\varepsilon_3)][1-f(\varepsilon_4)] \\ & \times \Big(\chi^+\left(f(\varepsilon_3)+f(\varepsilon_4)\right) \\ & \quad + \chi^-\left([1-f(\varepsilon_1)]+[1-f(\varepsilon_2)]\right)\Big)\,. \end{aligned} \tag{B.34}$$

The factors of $\chi^\pm$, which account for the difference in density of states of the two Rashba bands at their respective Fermi energies, can be approximated as 1, and we will therefore neglect them in the following.

Using the expressions for C and $\frac{\mathrm{d}C}{\mathrm{d}t}$, we want to find the relaxation-time constant τ in the following relaxation time approximation:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -\frac{C - C_{\text{eq.}}}{\tau}\,. \tag{B.35}$$

This yields

$$\frac{1}{\tau_{\text{ci}}} = \frac{2}{\mathscr{V}} \sum_{\mathbf{k}_1\mathbf{k}_2} \Phi^0(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^+_{\mathbf{k}_2}) \times w^{\text{car-imp}}_{(\mathbf{k}_1-)\to(\mathbf{k}_2+)} \bigg/ \frac{1}{\mathscr{V}} \sum_{\mathbf{k}_1} \Phi^0(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^+_{\mathbf{k}_1}) \tag{B.36}$$

and

$$\frac{1}{\tau_{\mathrm{cc}}} = \frac{2}{\mathscr{V}} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} \Phi^1(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^-_{\mathbf{k}_2}, \varepsilon^+_{\mathbf{k}_3}, \varepsilon^+_{\mathbf{k}_4}) \times w^{\mathrm{car-car}}_{(\mathbf{k}_1\mathbf{k}_2-)\to(\mathbf{k}_3\mathbf{k}_4+)} \Bigg/ \frac{1}{\mathscr{V}} \sum_{\mathbf{k}_1} \Phi^0(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^+_{\mathbf{k}_1}) \,. \tag{B.37}$$

Fermi's Golden Rule

The probability amplitudes in Eqs. (B.36) and (B.37) are given by Fermi's Golden Rule,

$$w_{(\mathbf{k}_3\mathbf{k}_4\mp)\to(\mathbf{k}_1\mathbf{k}_2\pm)} = \frac{2\pi}{\hbar} \, |\langle\Psi_{\mathrm{final}}| \, U \, |\Psi_{\mathrm{init}}\rangle|^2 \; \times \; \delta(\varepsilon^\pm_{\mathbf{k}_1} + \varepsilon^\pm_{\mathbf{k}_2} - \varepsilon^\mp_{\mathbf{k}_3} - \varepsilon^\mp_{\mathbf{k}_4}) \,, \tag{B.38}$$

where the subscripts init and fin refer to the initial and final states of the transition, respectively. We now have to find the matrix element for the transitions for carrier-impurity and carrier-carrier processes.

Initial and final states are taking the forms

$$\begin{aligned} \left|\Psi^{\mathrm{ci}}_{\mathrm{init}}\right\rangle &= |\mathbf{k}_1, -\rangle \,, & \left|\Psi^{\mathrm{ci}}_{\mathrm{fin}}\right\rangle &= |\mathbf{k}_2, +\rangle \,, \\ |\Psi^{\mathrm{cc}}_{\mathrm{init}}\rangle &= |\mathbf{k}_1, -\rangle \, |\mathbf{k}_2, -\rangle \,, & |\Psi^{\mathrm{cc}}_{\mathrm{fin}}\rangle &= |\mathbf{k}_3, +\rangle \, |\mathbf{k}_4, +\rangle \,. \end{aligned} \tag{B.39}$$

We note that a full treatment of this problem would envolve an antisymmetrisation of the two-particle wave-function, which would result in exchange-interaction terms in the matrix element. This however makes the integration that follows further below intractable. We also do not expect this to have a big impact onto the end result as exchange interaction only matters when a process and its exchange process are of similar strength, which is not true for the parts of the phase space that contribute dominantly to the scattering.

Each single-particle state can be written as a superposition of the Pauli matrix σ_Z eigenstates [129] as follows:

$$|\mathbf{k}, +\rangle = \frac{1}{\sqrt{2}} \left(|\mathbf{k}, \uparrow\rangle - \mathrm{i}e^{\mathrm{i}\theta_{\mathbf{k}}} \, |\mathbf{k}, \downarrow\rangle \right) \,, \tag{B.40}$$

$$|\mathbf{k}, -\rangle = \frac{1}{\sqrt{2}} \left(-\mathrm{i}e^{-\mathrm{i}\theta_{\mathbf{k}}} \, |\mathbf{k}, \uparrow\rangle + |\mathbf{k}, \downarrow\rangle \right) \,. \tag{B.41}$$

Because the component of U in spin space is the identity, its matrix element, $\langle \mathbf{k}', + | U | \mathbf{k}, - \rangle$, can be written as

$$\frac{1}{2}\Big(\langle \mathbf{k}',\uparrow| + \mathrm{i}\mathrm{e}^{-\mathrm{i}\theta_{\mathbf{k}'}}\langle \mathbf{k}',\downarrow|\Big)U\Big(-\mathrm{i}\mathrm{e}^{-\mathrm{i}\theta_{\mathbf{k}}}|\mathbf{k},\uparrow\rangle + |\mathbf{k},\downarrow\rangle\Big) \tag{B.42}$$

$$= \frac{1}{2}\Big(-\mathrm{i}\mathrm{e}^{-\mathrm{i}\theta_{\mathbf{k}'}} + \mathrm{i}\mathrm{e}^{-\mathrm{i}\theta_{\mathbf{k}}}\Big)\langle \mathbf{k}'| U |\mathbf{k}\rangle \tag{B.43}$$

$$= \frac{1}{2\mathrm{i}}\mathrm{e}^{-\frac{\mathrm{i}}{2}\left(\theta_{\mathbf{k}}+\theta_{\mathbf{k}'}\right)}\left(\mathrm{e}^{\frac{\mathrm{i}}{2}\left(\theta_{\mathbf{k}}-\theta_{\mathbf{k}'}\right)} - \mathrm{e}^{-\frac{\mathrm{i}}{2}\left(\theta_{\mathbf{k}}-\theta_{\mathbf{k}'}\right)}\right)\langle \mathbf{k}'| U |\mathbf{k}\rangle \tag{B.44}$$

$$= \mathrm{e}^{-\frac{\mathrm{i}}{2}\left(\theta_{\mathbf{k}}+\theta_{\mathbf{k}'}\right)}\sin\left(\theta_{\mathbf{k}}-\theta_{\mathbf{k}'}/2\right)\langle \mathbf{k}'| U |\mathbf{k}\rangle\,. \tag{B.45}$$

We set $\mathbf{p} = \mathbf{k} - \mathbf{k}'$ and hence $\langle \mathbf{k}'| U |\mathbf{k}\rangle = U_{\mathbf{p}}/\mathscr{V}$, where $\mathscr{V}$ is the area of the system and $U_{\mathbf{p}}$ is the Fourier transform of the Coulomb potential [119],

$$U_{\mathbf{p}} = U_p = \frac{2\pi e_0^2}{p + k_{\mathrm{S}}} \tag{B.46}$$

with $e_0^2 = e^2/4\pi\kappa\varepsilon_0$, where κ is the effective background lattice dielectric constant and k_{S} is the Thomas-Fermi screening momentum.

B.3.2 Impurity scattering

Next, we evaluate the time constant for carrier-impurity scattering, which is easy to compute as carrier-impurity scattering is a single-particle process that conserves energy.

Following the general considerations from the previous section, we find that $w^{\mathrm{car-imp}}_{(\mathbf{k}_1-)\to(\mathbf{k}_2+)}$ is equal to

$$\frac{2\pi}{\hbar\mathscr{V}^2}\sin^2\left(\theta_{\mathbf{k}_1}-\theta_{\mathbf{k}_2}/2\right)U^2_{|\mathbf{k}_1-\mathbf{k}_2|}\,\delta(\varepsilon^-_{\mathbf{k}_1} - \varepsilon^+_{\mathbf{k}_2})\,, \tag{B.47}$$

which we insert into Eq. (B.36) to obtain an expression for the time constant. We convert the sums to integrals to find

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{4\pi}{\hbar\mathscr{V}}\iint\frac{\mathrm{d}\mathbf{k}_1\mathrm{d}\mathbf{k}_2}{(2\pi)^4}\,\Phi^0(\varepsilon^-_{\mathbf{k}_1},\varepsilon^+_{\mathbf{k}_2})\,\delta(\varepsilon^-_{\mathbf{k}_1} - \varepsilon^+_{\mathbf{k}_2})\times$$
$$U^2_{|\mathbf{k}_1-\mathbf{k}_2|}\sin^2\left(\theta_{\mathbf{k}_1}-\theta_{\mathbf{k}_2}/2\right)\Bigg/\int\frac{\mathrm{d}\mathbf{k}_1}{(2\pi)^2}\,\Phi^0(\varepsilon^-_{\mathbf{k}_1},\varepsilon^+_{\mathbf{k}_1})\,. \tag{B.48}$$

We use the variable transformation $k_i\mathrm{d}k_i = \frac{2m}{\hbar^2}\,\mathrm{d}\varepsilon^{\pm}_i\,\frac{1}{2}\left(1 \pm \sqrt{E_{\mathrm{R}}/\varepsilon_i}\right)$ (for $i \in \{1,2\}$). Similar to our previous approximation of assuming $\chi^{\pm} \approx 1$, we drop the correction in the brackets, as this only accounts for a small correction stemming from the different densities of states of the two Rashba bands. Furthermore, we take the limit $T \to 0$, in which case $\Phi^0(\varepsilon_1, \varepsilon_2) \to \beta^{-1}(\delta(\varepsilon_1 - \mu) + \delta(\varepsilon_2 - \mu))$, to carry out the integrations over ε_1 and ε_2.

This yields for the denominator

$$\int \frac{\mathrm{d}\mathbf{k}_1}{(2\pi)^2} \Phi^0(\varepsilon^-_{\mathbf{k}_1}, \varepsilon^+_{\mathbf{k}_1}) = \frac{1}{2\pi} \frac{2m}{\hbar^2} \beta^{-1} . \tag{B.49}$$

Using this and carrying out the integration over ε_1 and ε_2 in the numerator, the result for timescale can be expressed as

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{1}{2\pi\hbar\mathcal{V}} \frac{2m}{\hbar^2} \int \mathrm{d}\theta_{\mathbf{k}_1,\mathbf{k}_2} U^2_{|\mathbf{k}_1 - \mathbf{k}_2|} \sin^2\left({\theta_{\mathbf{k}_1} - \theta_{\mathbf{k}_2}}/{2}\right) , \tag{B.50}$$

where $\theta_{\mathbf{k}_1,\mathbf{k}_2}$ is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$. By writing $|\mathbf{k}_{1/2}| = k_{\mathrm{F}}^{\mathrm{eq.}\pm} = k_{\mathrm{F}}^0 \mp k_{\mathrm{R}}$, we can derive that $|\mathbf{k}_1 - \mathbf{k}_2|^2 = 4(k_{\mathrm{F}}^0)^2 \sin^2(\theta/2) + 4k_{\mathrm{R}}^2 \cos^2(\theta/2) \approx 4(k_{\mathrm{F}}^0)^2 \sin^2(\theta/2) + 4k_{\mathrm{R}}^2$. Using this, we can write

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{1}{\tau_{\mathrm{ci},0}} \int_0^\pi \mathrm{d}\theta \frac{\sin^2(\theta/2)}{\left(2\sqrt{\sin^2(\theta/2) + \left(\frac{k_{\mathrm{R}}}{k_{\mathrm{F}}^0}\right)^2} + \frac{k_{\mathrm{S}}}{k_{\mathrm{F}}^0}\right)^2}, \tag{B.51}$$

where we have also multiplied the result with the number of impurity sites in the system N_{i} and introduced $n_{\mathrm{i}} = N_{\mathrm{i}}/\mathcal{V}$ and $\tau_{\mathrm{ci},0}^{-1} = 8\pi m n_{\mathrm{i}} e_0^4 / \hbar^3 (k_{\mathrm{F}}^0)^2$. In the last expression, we can observe that k_{R} appears in a similar way to the screening momentum k_{S}. Because, in strongly Rashba-coupled systems, k_{R} is several orders of magnitude larger than k_{S}, this therefore serves to enhance the screening of the Coulomb potential. More importantly, in the case without Rashba coupling, it is $k_{\mathrm{R}} = 0$ and the spinor-overlap matrix element vanishes, and we therefore find

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{1}{\tau_{\mathrm{ci},0}} \int_0^\pi \mathrm{d}\theta \frac{1}{\left(2\sin(\theta/2) + \frac{k_{\mathrm{S}}}{k_{\mathrm{F}}^0}\right)^2} \approx \frac{1}{\tau_{\mathrm{ci},0}} \frac{k_{\mathrm{F}}^0}{k_{\mathrm{S}}}, \tag{B.52}$$

whereas if we retain k_{R} and include the helical spin structure and use the fact that $k_{\mathrm{S}} \ll k_{\mathrm{R}}$, we instead find

$$\frac{1}{\tau_{\mathrm{ci}}} = \frac{1}{\tau_{\mathrm{ci},0}} \int_0^\pi \mathrm{d}\theta \frac{1}{4} \frac{\sin^2(\theta/2)}{\sin^2(\theta/2) + \left(\frac{k_{\mathrm{R}}}{k_{\mathrm{F}}^0}\right)^2} \approx \frac{1}{\tau_{\mathrm{ci},0}} \frac{\pi}{4}. \tag{B.53}$$

This results in an overall suppression factor of $\approx k_{\mathrm{S}}/k_{\mathrm{F}}^0$. We use Lindhard theory [86], which gives $k_{\mathrm{S}} = \frac{e^2 m}{\hbar^2}$, and using the effective mass $m \approx 0.02 m_{\mathrm{e}}$ for GeTe (m_e being the electron mass), we find ${}^{k_{\mathrm{S}}}/{}_{k_{\mathrm{F}}^{0}} \approx 23.7$, as reported in Sec. 6.3. While this suppression is not as dramatic as in the phonon-scattering and inter-carrier cases, we note that to obtain the effective rate for carrier-impurity scattering, this suppression factor must be multiplied by a density of impurities. Therefore, in clean, undoped samples, in which the concentration of unintentional dopants is negligibly small, we expect this not to be of importance.

B.3.3 Inter-carrier scattering

We now continue with examining the carrier-carrier contribution to relaxation, which is harder and requires significantly more work.

Inserting Fermi's Golden Rule and converting sums to integrals

Using the result in Eqs. (B.45) and (B.46), the matrix element squared, $|\langle\Psi_{\mathrm{final}}|\,U\,|\Psi_{\mathrm{init}}\rangle\,|^2$, can be computed to be

$$\frac{U_q^2}{\mathcal{V}^2}\,\sin^2\left({}^{\theta_{\mathbf{k}_1,\mathbf{k}_1+\mathbf{q}}}/{}_2\right)\sin^2\left({}^{\theta_{\mathbf{k}_2,\mathbf{k}_2-\mathbf{q}}}/{}_2\right)\;\delta_{\mathbf{k}_4,\mathbf{k}_2-\mathbf{q}}\,, \tag{B.54}$$

where we have defined $\mathbf{q}=\mathbf{k}_3-\mathbf{k}_1$, and where $\theta_{\mathbf{k}_1,\mathbf{k}_1+\mathbf{q}}$ ($\theta_{\mathbf{k}_2,\mathbf{k}_2-\mathbf{q}}$) is the angle between initial wavevector $\mathbf{k}_1$ ($\mathbf{k}_2$) and final wavevector $\mathbf{k}_1+\mathbf{q}$ ($\mathbf{k}_2-\mathbf{q}$). Using Fermi's Golden Rule in Eq. (B.38), we can insert this into the expression for the relaxation-time constant in Eq. (B.37) and convert the sums over momentum space to continuous integrals. This results in

$$\begin{aligned}\frac{1}{\tau_{\mathrm{cc}}} =& \frac{4\pi}{\hbar}\iiint\frac{\mathrm{d}\mathbf{k}_1\mathrm{d}\mathbf{k}_2\mathrm{d}\mathbf{q}}{(2\pi)^6}\,U_q^2\,\Phi^1(\varepsilon^-_{\mathbf{k}_1},\varepsilon^-_{\mathbf{k}_2},\varepsilon^+_{\mathbf{k}_3},\varepsilon^+_{\mathbf{k}_4})\\ &\times\;\sin^2\left({}^{\theta_{\mathbf{k}_1,\mathbf{k}_1+\mathbf{q}}}/{}_2\right)\sin^2\left({}^{\theta_{\mathbf{k}_2,\mathbf{k}_2-\mathbf{q}}}/{}_2\right)\\ &\times\;\delta(\varepsilon^-_{\mathbf{k}_1}+\varepsilon^-_{\mathbf{k}_2}-\varepsilon^+_{\mathbf{k}_1+\mathbf{q}}-\varepsilon^+_{\mathbf{k}_2-\mathbf{q}})\Bigg/\int\frac{\mathrm{d}\mathbf{k}_1}{(2\pi)^2}\,\Phi^0(\varepsilon^+_{\mathbf{k}_1},\varepsilon^-_{\mathbf{k}_1})\,.\end{aligned} \tag{B.55}$$

Defining variable transformations

What follows is a series of variable transformations. As in the impurity case, we use $k_i\,\mathrm{d}k_i=\frac{2m}{\hbar^2}\,\mathrm{d}\varepsilon_i\,\frac{1}{2}\,(1\pm\sqrt{{}^{E_{\mathrm{R}}}/{}_{\varepsilon_i}})$ ($i\in\{1,2\}$), and the denominator again becomes $\frac{1}{2\pi}\,\frac{2m}{\hbar^2}\,\beta^{-1}$. Next, we follow an approach by Lawrence and Wilkins [78] to write the integrals over $\mathbf{k}_1$ and $\mathbf{k}_2$ in terms of the variables $\varepsilon_1=\varepsilon^-_{\mathbf{k}_1}$, $\varepsilon_2=\varepsilon^-_{\mathbf{k}_2}$, $\varepsilon_p=\varepsilon^+_{\mathbf{k}_1+\mathbf{q}}$, and $\varepsilon_{p'}=\varepsilon^+_{\mathbf{k}_2-\mathbf{q}}$.

First, we find an expression for the angle between $\mathbf{k}_1$ and $\mathbf{q}$, which we call $\theta_{\mathbf{k}_1,\mathbf{q}}$, as a function of the new variables. It is $\varepsilon_1={}^{\hbar^2}/{}_{2m}\,(k_1-k_{\mathrm{R}})^2$ and $\varepsilon_p={}^{\hbar^2}/{}_{2m}\,(|\mathbf{k}_1+\mathbf{q}|+k_{\mathrm{R}})^2$, and by writing $(\mathbf{k}_1+\mathbf{q})^2 = k_1^2+q^2+2k_1q\cos\theta_{\mathbf{k}_1,\mathbf{q}}$, we find that

$$\cos\theta_{\mathbf{k}_1,\mathbf{q}} = \frac{|\mathbf{k}_1+\mathbf{q}|^2-k_1^2-q^2}{2k_1q} \tag{B.56}$$

$$= -\frac{\varepsilon_1-\varepsilon_p+\varepsilon_q+2\sqrt{E_\mathrm{R}}(\sqrt{\varepsilon_p}+\sqrt{\varepsilon_1})}{2\sqrt{\varepsilon_q}(\sqrt{\varepsilon_1}+\sqrt{E_\mathrm{R}})} \tag{B.57}$$

and accordingly for the angle $\theta_{\mathbf{k}_2,\mathbf{q}}$ between $\mathbf{k}_2$ and $\mathbf{q}$ and variable $\varepsilon_{p'}$. Therefore, we find the derivative of the angles with respect ε_p and $\varepsilon_{p'}$ to be

$$\frac{\partial\theta_{\mathbf{k}_1,\mathbf{q}}}{\partial\varepsilon_p} = -\Omega(\varepsilon_1,\varepsilon_p,\varepsilon_q)\left(1-\sqrt{\frac{E_\mathrm{R}}{\varepsilon_p}}\right), \tag{B.58}$$

$$\frac{\partial\theta_{\mathbf{k}_2,\mathbf{q}}}{\partial\varepsilon_{p'}} = \Omega(\varepsilon_2,\varepsilon_{p'},\varepsilon_q)\left(1-\sqrt{\frac{E_\mathrm{R}}{\varepsilon_{p'}}}\right). \tag{B.59}$$

where $\Omega(\varepsilon_1,\varepsilon_p,\varepsilon_q)$ is defined as

$$\frac{1}{\sqrt{4\varepsilon_q\left(\sqrt{\varepsilon_1}+\sqrt{E_\mathrm{R}}\right)^2-\left(\varepsilon_1-\varepsilon_p+\varepsilon_q+2\sqrt{E_\mathrm{R}}(\sqrt{\varepsilon_p}+\sqrt{\varepsilon_1})\right)^2}}. \tag{B.60}$$

Using this result, we can find the Jacobian determinants for the transformation from $\mathbf{k}_1$ to $(\varepsilon_1,\varepsilon_p)$ and $\mathbf{k}_2$ to $(\varepsilon_2,\varepsilon_{p'})$. We find that $k_1\mathrm{d}k_1\,\mathrm{d}\theta_{\mathbf{k}_1,\mathbf{q}}$ equals to

$$\mathrm{d}\varepsilon_1\mathrm{d}\varepsilon_p\frac{2m}{\hbar^2}\,\Omega(\varepsilon_1,\varepsilon_p,\varepsilon_q)\left(1+\sqrt{\frac{E_\mathrm{R}}{\varepsilon_1}}\right)\left(1-\sqrt{\frac{E_\mathrm{R}}{\varepsilon_p}}\right) \tag{B.61}$$

and the same expression for $k_2\mathrm{d}k_2\,\mathrm{d}\theta_{\mathbf{k}_2,\mathbf{q}}$ under exchange of $\varepsilon_1\mapsto\varepsilon_2$ and $\varepsilon_p\mapsto\varepsilon_{p'}$. As was done previously, we drop the last two terms in the brackets as these small corrections account for the difference in the densities of states. We have included an extra factor of 2 because the cos is only uniquely defined on the $[0,\pi]$ interval and so accounts for only half of the integral that we want to calculate. The boundaries of the ε_p integral will be such that the momenta of $\mathbf{k}_1$ and $\mathbf{q}$ either align or antialign, and so we find

$$\varepsilon_p^{\mathrm{max/min}} = \frac{\hbar^2}{2m}(|k_1\pm q|+k_\mathrm{R})^2, \tag{B.62}$$

which can be rewritten as

$$\varepsilon_p^{\max} = (\sqrt{\varepsilon_1} + \sqrt{\varepsilon_q} + 2\sqrt{E_{\mathrm{R}}})^2, \tag{B.63}$$

$$\varepsilon_p^{\min} = \begin{cases} (\sqrt{\varepsilon_1} - \sqrt{\varepsilon_q} + 2\sqrt{E_{\mathrm{R}}})^2, & \text{for } \varepsilon_q < \varepsilon_1 , \\ (\sqrt{\varepsilon_1} - \sqrt{\varepsilon_q})^2, & \text{for } \varepsilon_q > \varepsilon_1 . \end{cases} \tag{B.64}$$

and the respective result for $\varepsilon_{p'}$ with $\varepsilon_1 \mapsto \varepsilon_2$. Finally, it is $q\,\mathrm{d}q = {}^1\!/\!{}_2\,({}^{2m}\!/\!{}_{\hbar^2})\,\mathrm{d}\varepsilon_q$ with $\varepsilon_q = {}^{\hbar^2 q^2}\!/\!{}_{2m}$. The integration over the remaining free angle results in an additional factor of 2π. Using this and Eq. (B.61), we can rewrite Eq. (B.55) as

$$\begin{aligned} \frac{1}{\tau_{\mathrm{cc}}} =& \frac{1}{2\pi\hbar}\left(\frac{2m}{\hbar^2}\right)^2 \beta \int_{-\infty}^{+\infty} \mathrm{d}\varepsilon_1 \int_{-\infty}^{+\infty} \mathrm{d}\varepsilon_2 \int_0^{\infty} \mathrm{d}\varepsilon_q \int_{\varepsilon_p^{\min}}^{\varepsilon_p^{\max}} \mathrm{d}\varepsilon_p \int_{\varepsilon_{p'}^{\min}}^{\varepsilon_{p'}^{\max}} \mathrm{d}\varepsilon_{p'} \\ & \times U_{\varepsilon_q}^2\, \Phi^1(\varepsilon_1, \varepsilon_2, \varepsilon_p, \varepsilon_{p'})\, S(\varepsilon_1, \varepsilon_p) S(\varepsilon_2, \varepsilon_{p'}) \\ & \times \Omega(\varepsilon_1, \varepsilon_p, \varepsilon_q)\Omega(\varepsilon_2, \varepsilon_{p'}, \varepsilon_q)\, \delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_p - \varepsilon_{p'}), \end{aligned} \tag{B.65}$$

where the transformed Coulomb potential, U_{ε_q}, and spinor-overlap matrix elements, $S(\varepsilon_1, \varepsilon_p)$ and $S(\varepsilon_2, \varepsilon_{p'})$, will be provided in the next section.

Applying variable transformations to integrand

Having defined suitable variable transformations and having calculated their Jacobian determinants, we will proceed by applying the transformation to the integrand.

The Coulomb potential can easily be rewritten as

$$U_{\varepsilon_q} = \left(\frac{2m}{\hbar^2}\right)^{-1/2} \frac{2\pi e_0^2}{\sqrt{\varepsilon_q} + \sqrt{\varepsilon_S}}, \tag{B.66}$$

where we have defined $\varepsilon_S = {}^{\hbar^2 k_{\mathrm{S}}^2}\!/\!{}_{2m}$.

We can relate the angle between $\mathbf{k}_1$ and $\mathbf{k}_1 + \mathbf{q}$ ($\mathbf{k}_2$ and $\mathbf{k}_2 - \mathbf{q}$) to the angle between $\mathbf{k}_1$ ($\mathbf{k}_2$) and $\mathbf{q}$,

$$\cos\left(\theta_{\mathbf{k}_1, \mathbf{k}_1 + \mathbf{q}}\right) = \frac{k_1 + q\cos\left(\theta_{\mathbf{k}_1, \mathbf{q}}\right)}{|\mathbf{k}_1 + \mathbf{q}|} \tag{B.67}$$

and

$$\cos\left(\theta_{\mathbf{k}_2, \mathbf{k}_2 - \mathbf{q}}\right) = \frac{k_2 - q\cos\left(\theta_{\mathbf{k}_2, \mathbf{q}}\right)}{|\mathbf{k}_2 - \mathbf{q}|}, \tag{B.68}$$

and, using $\sin^2(\theta/2) = 1/2\,(1-\cos(\theta))$, we find

$$\begin{aligned}
\sin^2\left(\theta_{\mathbf{k}_1,\mathbf{k}_1+\mathbf{q}}/2\right) &= \frac{1}{2}\frac{|\mathbf{k}_1+\mathbf{q}|-k_1-q\cos\left(\theta_{\mathbf{k}_1,\mathbf{q}}\right)}{|\mathbf{k}_1+\mathbf{q}|} && \text{(B.69)}\\
&\overset{\text{(B.56)}}{=} \frac{1}{4}\frac{q^2-(k_1-|\mathbf{k}_1+\mathbf{q}|)^2}{k_1|\mathbf{k}_1+\mathbf{q}|}, && \text{(B.70)}\\
\sin^2\left(\theta_{\mathbf{k}_2,\mathbf{k}_2-\mathbf{q}}/2\right) &= \frac{1}{2}\frac{|\mathbf{k}_2-\mathbf{q}|-k_2+q\cos\left(\theta_{\mathbf{k}_2,\mathbf{q}}\right)}{|\mathbf{k}_2-\mathbf{q}|} && \text{(B.71)}\\
&= \frac{1}{4}\frac{q^2-(k_2-|\mathbf{k}_2-\mathbf{q}|)^2}{k_2|\mathbf{k}_2-\mathbf{q}|}. && \text{(B.72)}
\end{aligned}$$

Therefore, we can define the spinor-overlap function $S(\varepsilon_1,\varepsilon_p,\varepsilon_q)$ used in Eq. (B.65) as

$$\frac{1}{4}\frac{\varepsilon_q-(\sqrt{\varepsilon_1}-\sqrt{\varepsilon_p}+2\sqrt{E_\mathrm{R}})^2}{(\sqrt{\varepsilon_1}+\sqrt{E_\mathrm{R}})(\sqrt{\varepsilon_p}-\sqrt{E_\mathrm{R}})}. \tag{B.73}$$

Computing the integral

In this section, we will proceed by computing the integral in Eq. (B.65). This will yield an expression for the inter-carrier scattering relaxation timescale τ_cc.

We define $\varepsilon_\Delta = \varepsilon_p - \varepsilon_1$ and $\varepsilon_{\Delta'} = \varepsilon_{p'} - \varepsilon_2$ and write Eq. (B.65) as

$$\begin{aligned}
\frac{1}{\tau_\mathrm{cc}} =& \frac{\beta}{\tau_{cc,0}}\int_{-\infty}^{+\infty}\mathrm{d}\varepsilon_1\int_{-\infty}^{+\infty}\mathrm{d}\varepsilon_2\int_0^{\infty}\mathrm{d}\varepsilon_q\int_{\varepsilon_\Delta^\mathrm{min}}^{\varepsilon_\Delta^\mathrm{max}}\mathrm{d}\varepsilon_\Delta\int_{\varepsilon_\Delta^\mathrm{min}}^{\varepsilon_\Delta^\mathrm{max}}\mathrm{d}\varepsilon_\Delta\, u_{\varepsilon_q}^2 \\
&\times\Phi^1(\varepsilon_1,\varepsilon_2,\varepsilon_1{+}\varepsilon_\Delta,\varepsilon_2{+}\varepsilon_{\Delta'})\,S(\varepsilon_1,\varepsilon_1{+}\varepsilon_\Delta)S(\varepsilon_2,\varepsilon_2{+}\varepsilon_{\Delta'})\\
&\times\Omega(\varepsilon_1,\varepsilon_1+\varepsilon_\Delta,\varepsilon_q)\Omega(\varepsilon_2,\varepsilon_2+\varepsilon_{\Delta'},\varepsilon_q)\,\delta(\varepsilon_\Delta{+}\varepsilon_{\Delta'}),
\end{aligned}\tag{B.74}$$

where $u_{\varepsilon_q} = (\sqrt{\varepsilon_q}+\sqrt{\varepsilon_S})^{-1}$ and

$$\frac{1}{\tau_{cc,0}} = \frac{(2\pi e_0^2)^2}{2\pi\hbar}\frac{2m}{\hbar^2} = \frac{1}{8\pi}\frac{m}{\hbar^3}\frac{e^4}{\varepsilon_0^2\,\kappa^2}. \tag{B.75}$$

The boundaries of the integration over ε_Δ can be deduced from Eq. (B.63) and are given by

$$\varepsilon_\Delta^{\max/\min} = (2\sqrt{E_\mathrm{R}}\pm\sqrt{\varepsilon_q})^2+2\sqrt{\varepsilon_1}(2\sqrt{E_\mathrm{R}}\pm\sqrt{\varepsilon_q}) \tag{B.76}$$

for $\varepsilon_q < \varepsilon_1$ and

$$\varepsilon_{\Delta}^{\max} = (2\sqrt{E_{\mathrm{R}}} + \sqrt{\varepsilon_q})^2 + 2\sqrt{\varepsilon_1}(2\sqrt{E_{\mathrm{R}}} + \sqrt{\varepsilon_q})\,, \tag{B.77}$$

$$\varepsilon_{\Delta}^{\min} = \varepsilon_q - 2\sqrt{\varepsilon_1 \varepsilon_q} \tag{B.78}$$

for $\varepsilon_q > \varepsilon_1$ (and accordingly for $\varepsilon_{\Delta'}$).

In order for the δ function to give a contribution, it must be $\varepsilon_{\Delta}^{\max} + \varepsilon_{\Delta'}^{\max} > 0$ and $\varepsilon_{\Delta}^{\min} + \varepsilon_{\Delta'}^{\min} < 0$. The first condition is trivial and can be omitted. In the case $\varepsilon_q < \varepsilon_1$, the second condition requires that

$$\left(2\sqrt{E_{\mathrm{R}}} - \sqrt{\varepsilon_q}\right)^2 + \left(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}\right)\left(2\sqrt{E_{\mathrm{R}}} - \sqrt{\varepsilon_q}\right) < 0\,. \tag{B.79}$$

This necessitates that $\sqrt{\varepsilon_q} > 2\sqrt{E_{\mathrm{R}}}$, which is just the observation that $q > 2k_{\mathrm{R}}$ that we discuss in detail in the main text. When this condition is satisfied, the inequality is inverted when dividing by $\left(2\sqrt{E_{\mathrm{R}}} - \sqrt{\varepsilon_q}\right)$, and we find

$$\left(2\sqrt{E_{\mathrm{R}}} - \sqrt{\varepsilon_q}\right) + \left(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}\right) > 0\,, \tag{B.80}$$

which is always true for the case $\varepsilon_q < \varepsilon_1$ that we started with. In the case of $\varepsilon_q > \varepsilon_1$ (in which case we also assert that $\varepsilon_q > \varepsilon_2$), the condition $\varepsilon_{\Delta}^{\min} + \varepsilon_{\Delta'}^{\min} < 0$ requires that

$$\sqrt{\varepsilon_q} < \sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}\,. \tag{B.81}$$

We can now execute the integration over $\varepsilon_{\Delta'}$, which is equivalent to setting $\varepsilon_{\Delta'} = -\varepsilon_{\Delta}$ due to the δ function. The integration over ε_{Δ} will then run from $\varepsilon_{\min} = \max\left(\varepsilon_{\Delta}^{\min}, -\varepsilon_{\Delta'}^{\max}\right)$ to $\varepsilon_{\max} = \min\left(\varepsilon_{\Delta}^{\max}, -\varepsilon_{\Delta'}^{\min}\right)$. To find the min or max, we check whether $\varepsilon_{\Delta}^{\min/\max} + \varepsilon_{\Delta'}^{\max/\min} > 0$, which in the case of $\varepsilon_q < \varepsilon_1$ is equivalent to

$$4E_{\mathrm{R}} + \varepsilon_q + 2\sqrt{E_{\mathrm{R}}}(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}) \mp \sqrt{\varepsilon_q}(\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}) > 0\,. \tag{B.82}$$

Because $|\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}| \approx \sqrt{k_{\mathrm{B}}T} \ll \sqrt{E_{\mathrm{R}}} < \sqrt{\varepsilon_q}$, this is always true. Furthermore, in the case $\varepsilon_q > \varepsilon_1$ the above condition is equivalent to

$$(2\sqrt{E_{\mathrm{R}}} + \sqrt{\varepsilon_q})^2 + \varepsilon_q + 2\sqrt{\varepsilon_{2/1} E_{\mathrm{R}}} \mp 2\sqrt{\varepsilon_q}(\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}) > 0\,, \tag{B.83}$$

which is satisfied again because $|\sqrt{\varepsilon_1}-\sqrt{\varepsilon_2}| \ll \sqrt{\varepsilon_q}$. Therefore, $\varepsilon_{\min} = \varepsilon_\Delta^{\min}$ and $\varepsilon_{\max} = -\varepsilon_{\Delta'}^{\min}$. Using these results, we can now write

$$\begin{aligned}\frac{1}{\tau_{\mathrm{cc}}} =& \frac{\beta}{\tau_{\mathrm{cc},0}} \int \mathrm{d}\varepsilon_1 \int \mathrm{d}\varepsilon_2 \int_{4E_{\mathrm{R}}}^{(\sqrt{\varepsilon_1}+\sqrt{\varepsilon_2})^2} \mathrm{d}\varepsilon_q \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \mathrm{d}\varepsilon_\Delta \, u_{\varepsilon_q}^2 \\ &\times \Phi^1(\varepsilon_1,\varepsilon_2,\varepsilon_1+\varepsilon_\Delta,\varepsilon_2-\varepsilon_\Delta)\, S(\varepsilon_1,\varepsilon_1+\varepsilon_\Delta) S(\varepsilon_2,\varepsilon_2-\varepsilon_\Delta) \\ &\times \Omega(\varepsilon_1,\varepsilon_1+\varepsilon_\Delta,\varepsilon_q)\Omega(\varepsilon_2,\varepsilon_2-\varepsilon_\Delta,\varepsilon_q)\,.\end{aligned} \tag{B.84}$$

So far, all manipulations have been exact. We will continue with a number of approximations to be able to find a closed expression for the relaxation time. The spinor-overlap function $S(\varepsilon_1,\varepsilon_1+\varepsilon_\Delta,\varepsilon_q)$ takes the form

$$\frac{1}{4}\frac{\varepsilon_q - (\sqrt{\varepsilon_1}-\sqrt{\varepsilon_1+\varepsilon_\Delta}+2\sqrt{E_{\mathrm{R}}})^2}{(\sqrt{\varepsilon_1}+\sqrt{E_{\mathrm{R}}})(\sqrt{\varepsilon_1+\varepsilon_\Delta}-\sqrt{E_{\mathrm{R}}})}\,. \tag{B.85}$$

ε_Δ is of order $k_{\mathrm{B}}T$ and therefore much smaller than $\varepsilon_1 \approx \mu$ and E_{R}. We therefore expand the denominator in lowest order in ε_Δ and drop its dependence in the numerator, which gives

$$\frac{1}{4}\frac{\sqrt{\varepsilon_1}(\varepsilon_q-4E_{\mathrm{R}})+2\varepsilon_\Delta\sqrt{E_{\mathrm{R}}}}{\sqrt{\varepsilon_1}(\varepsilon_1-E_{\mathrm{R}})}\,. \tag{B.86}$$

Similarly, we approximate $\Omega(\varepsilon_1,\varepsilon_1+\varepsilon_\Delta,\varepsilon_q)$ (defined in Eq. (B.60)) as

$$\frac{1}{\sqrt{4\varepsilon_q\big(\sqrt{\varepsilon_1}+\sqrt{E_{\mathrm{R}}}\big)^2-\big(-\varepsilon_\Delta+\varepsilon_q+4\sqrt{\varepsilon_1 E_{\mathrm{R}}}\big)^2}}\,. \tag{B.87}$$

Next, we note that ε_1 and ε_2 are approximately equal to μ, and variations around this value are on the order of $k_{\mathrm{B}}T$. We can therefore neglect these variations and replace ε_1 and ε_2 by μ everywhere except for in the Fermi distributions in Φ^1. Consequently, we can perform the integrations over ε_1 and ε_2. We use the following expression for $\Phi^1(\varepsilon_1,\varepsilon_2,\varepsilon_1+\varepsilon_\Delta,\varepsilon_2-\varepsilon_\Delta)$:

$$\begin{aligned}& f(\varepsilon_1+\varepsilon_\Delta)f(\varepsilon_2-\varepsilon_\Delta)[1-f(\varepsilon_1)][1-f(\varepsilon_2)] \\ &\times\Big([1-f(\varepsilon_1+\varepsilon_\Delta)]+[1-f(\varepsilon_2-\varepsilon_\Delta)]+f(\varepsilon_1)+f(\varepsilon_2)\Big) \\ &+f(\varepsilon_1)f(\varepsilon_2)[1-f(\varepsilon_1+\varepsilon_\Delta)][1-f(\varepsilon_2-\varepsilon_\Delta)] \\ &\times\Big(f(\varepsilon_1+\varepsilon_\Delta)+f(\varepsilon_2-\varepsilon_\Delta)+[1-f(\varepsilon_1)]+[1-f(\varepsilon_2)]\Big)\,,\end{aligned} \tag{B.88}$$

where the second term gives the same contribution as the first, as can be seen by renaming $\varepsilon_1+\varepsilon_\Delta \mapsto \varepsilon_2$ and $\varepsilon_2-\varepsilon_\Delta \mapsto \varepsilon_1$. Finally, a straight-forward integration over ε_1 and ε_2 gives

$$\int \varepsilon_1 \int \varepsilon_2 \, \Phi^1(\varepsilon_1,\varepsilon_2,\varepsilon_1+\varepsilon_\Delta,\varepsilon_2-\varepsilon_\Delta) = \frac{4\,\varepsilon_\Delta^2\, e^{\beta\varepsilon_\Delta}}{(e^{\beta\varepsilon_\Delta}-1)^2}\,. \tag{B.89}$$

This allows us to write

$$\begin{aligned}\frac{1}{\tau_{\mathrm{cc}}} = {}& \frac{(k_{\mathrm{B}}T)^2}{\tau_{\mathrm{cc},0}} \int_{4E_{\mathrm{R}}}^{4\mu} \mathrm{d}\varepsilon_q \, u_{\varepsilon_q}^2 \int_{-\varepsilon_{\max}}^{\varepsilon_{\max}} \beta \, \mathrm{d}\varepsilon_\Delta \\ & \times \frac{4(\beta\varepsilon_\Delta)^2 e^{\beta\varepsilon_\Delta}}{(e^{\beta\varepsilon_\Delta}-1)^2} \frac{1}{16} \frac{\mu(\varepsilon_q-4E_{\mathrm{R}})^2-4\varepsilon_\Delta^2 E_{\mathrm{R}}}{\mu(\mu-E_{\mathrm{R}})^2} \\ & \times \frac{1}{\sqrt{4\varepsilon_q(\sqrt{\mu}+\sqrt{E_{\mathrm{R}}})^2-(\varepsilon_q+4\sqrt{\mu E_{\mathrm{R}}}+\varepsilon_\Delta)^2}} \\ & \times \frac{1}{\sqrt{4\varepsilon_q(\sqrt{\mu}+\sqrt{E_{\mathrm{R}}})^2-(\varepsilon_q+4\sqrt{\mu E_{\mathrm{R}}}-\varepsilon_\Delta)^2}}\,.\end{aligned} \tag{B.90}$$

We use the lowest-order expansion in ε_Δ and use the substitution $x = \beta\varepsilon_\Delta$ to find

$$\frac{1}{\tau_{\mathrm{cc}}} = \frac{1}{\tau_{cc,0}} \frac{(k_{\mathrm{B}}T)^2}{(\mu-E_{\mathrm{R}})^2} \int_{4E_{\mathrm{R}}}^{4\mu} \mathrm{d}\varepsilon_q \, \frac{\varepsilon_q-4E_{\mathrm{R}}}{\varepsilon_q(4\mu-\varepsilon_q)} \, \gamma(\beta\varepsilon_{\max})\,, \tag{B.91}$$

where we've set $\varepsilon_S = 0$ (because its effect on screening can be neglected), and where we have defined

$$\gamma(y) = \frac{1}{2}\int_0^y \mathrm{d}x \, \frac{x^2 e^x}{(e^x-1)^2}\,. \tag{B.92}$$

Where $\varepsilon_{\max} = \sqrt{\varepsilon_q}(\sqrt{4\mu}-\sqrt{\varepsilon_q}) \gg k_{\mathrm{B}}T$, we can replace $\gamma(\beta\varepsilon_{\max})$ with $\gamma(\infty) = \frac{\pi^2}{6}$. This holds everywhere except for ε_q close to 4μ, which is the upper limit of the ε_q integral and also where the rest of the integrand is logarithmically divergent (due to the $(4\mu-\varepsilon_q)^{-1}$ contribution). We therefore have to split the ε_q integration up into one part close to $\varepsilon_q = 4\mu$ and one part further away from this point. As shown in Fig. B.1, we can approximate $\gamma(y)$ as

$$\gamma(y) = \begin{cases} \frac{\pi^2}{6}\,, & \text{if } y > \frac{\pi^2}{3}\,, \\ \frac{y}{2}\,, & \text{if } y \leq \frac{\pi^2}{3}\,. \end{cases} \tag{B.93}$$

We solve $\beta\varepsilon_{\max} = \frac{\pi^2}{6}$ for ε_q to find

$$\varepsilon_q = 2\mu\left(1 - \frac{\pi^2}{6}\frac{k_\mathrm{B}T}{\mu} + \sqrt{1 - \frac{\pi^2}{3}\frac{k_\mathrm{B}T}{\mu}}\right) \tag{B.94}$$

$$\approx 4\mu\left(1 - \frac{\pi^2}{6}\frac{k_\mathrm{B}T}{\mu}\right) =: \varepsilon_q^* \,. \tag{B.95}$$

We can therefore write

$$\frac{1}{\tau_\mathrm{cc}} = \int_{4E_\mathrm{R}}^{4\mu} \mathrm{d}\varepsilon_q \, \frac{\varepsilon_q - 4E_\mathrm{R}}{\varepsilon_q(4\mu - \varepsilon_q)} \, \gamma(\beta\varepsilon_{\max}) \tag{B.96}$$

$$\begin{aligned} &= \frac{\pi^2}{6}\int_{4E_\mathrm{R}}^{\varepsilon_q^*} \mathrm{d}\varepsilon_q \, \frac{\varepsilon_q - 4E_\mathrm{R}}{\varepsilon_q(4\mu - \varepsilon_q)} \\ &+ \frac{1}{2}\int_{\varepsilon_q^*}^{4\mu} \mathrm{d}\varepsilon_q \, \frac{\varepsilon_q - 4E_\mathrm{R}}{\sqrt{\varepsilon_q}(\sqrt{4\mu} + \sqrt{\varepsilon_q})} \,, \end{aligned} \tag{B.97}$$

and since $\varepsilon_q^* \approx 4\mu$ we can set $\varepsilon_q = 4\mu$ in the second integral, which results in

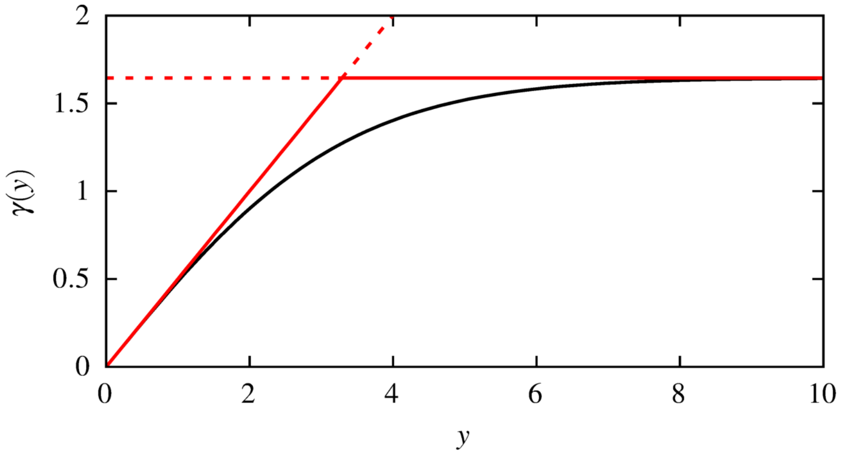
$$\begin{aligned} \frac{1}{\tau_\mathrm{cc}} &= \frac{\pi^2}{6}\left[-\frac{E_\mathrm{R}}{\mu}\log(\varepsilon_q) - \frac{(\mu - E_\mathrm{R})\log(4\mu - \varepsilon_q)}{\mu}\right]_{\varepsilon_q = 4E_\mathrm{R}}^{\varepsilon_q = \varepsilon_q^*} \\ &+ \frac{4\mu}{2}\frac{\pi^2}{6}\frac{k_\mathrm{B}T}{\mu}\frac{4\mu - 4E_\mathrm{R}}{2\sqrt{\mu}\,(\sqrt{4\mu} + \sqrt{4\mu})} \end{aligned} \tag{B.98}$$

$$\begin{aligned} &\approx \frac{\pi^2}{6}\left(1 - \frac{E_\mathrm{R}}{\mu}\left(1 - \log\left(\frac{E_\mathrm{R}}{\mu}\right)\right)\right. \\ &\quad \left. - \left(1 - \frac{E_\mathrm{R}}{\mu}\right)\log\left(\frac{\pi^2}{6}\frac{k_\mathrm{B}T}{\mu - E_\mathrm{R}}\right)\right) , \end{aligned} \tag{B.99}$$

which we define as $\rho\left(\frac{E_\mathrm{R}}{\mu}, \frac{k_\mathrm{B}T}{\mu - E_\mathrm{R}}\right)$. We can therefore write

$$\frac{1}{\tau_\mathrm{cc}} = \frac{1}{\tau_{\mathrm{cc},0}}\frac{(k_\mathrm{B}T)^2}{(\mu - E_\mathrm{R})^2} \times \rho\left(\frac{E_\mathrm{R}}{\mu}, \frac{k_\mathrm{B}T}{\mu - E_\mathrm{R}}\right) . \tag{B.100}$$

Using Eq. (B.75) and $\kappa = 10$, we find $\tau_{\mathrm{cc},0} = 10^{-11}\,\mathrm{s}$. For thin films of GeTe with $\mu - E_\mathrm{R} = 0.1\,\mathrm{eV}$ [82, 99] and $\frac{E_\mathrm{R}}{\mu} \approx 0.5$, which gives $\rho\left(\frac{E_\mathrm{R}}{\mu}, \frac{k_\mathrm{B}T}{\mu - E_\mathrm{R}}\right) \approx 10$, and Eq. (B.100) yields $\tau_\mathrm{cc} \approx 1\,\mu\mathrm{s}$ at $T = 100\,\mathrm{mK}$ as reported in Sec. 6.3.



Reprinted with permission from Ref. [143], P.C. Verpoort et al., J. Phys.: Condens. Matter 32, 355704 (2020). *Licensed under Creative Commons Attribution 4.0.*

Fig. B.1 Black curve: $\gamma(y)$, as defined in Eq. (B.92). Red curve: approximation of $\gamma(y)$, as defined in Eq. (B.93).

B.4 The role of spin-flips

While the phonon-mediated transitions are prohibited independently of spin due to energy-momentum conservation, we made explicit use of the anti-alignment of spin eigenstates for low-q inter-band transitions mediated by the Coulomb interaction in Sec. 6.3 and assumed the absence of spin-flip processes. It is however well known that the spin of a carrier can be flipped through various different mechanisms, and it is therefore worthwhile asking whether any of these can enhance the inter-band scattering rate and ultimately invalidate the arguments presented in Sec. 6.3.

The spin of a carrier can either be flipped through interaction with magnetic impurities, which can be disregarded in high-purity samples, or alternatively through strong spin-orbit coupling, which is a crucial ingredient of our work. One may therefore ask whether the spin-orbit coupling of the Rashba system can induce spin-flips that will lead to fast inter-band equilibration.

First we note that it is important not to confuse the spin-relaxation timescale with the likelihood for spin-reversing processes to occur. We expect spin relaxation to continue to occur on very short timescales in our system, as spin is not a conserved quantity in each individual Rashba band and scattering events that conserve the chiral index will occur at a high rate. To relax the non-equilibrium carrier configuration that we study, spin-reversing inter-band scattering processes are required, and there is no direct relation between their rate and the spin-relaxation timescale.

Next, we note that while spin-orbit coupling by itself leads to spin mixing of electronic eigenstates, no spin-flips can occur without a scattering process that mediates it. We have identified all relevant scattering processes for the relaxation of the depicted non-equilibrium state in Sec. 6.1, and we would now have to continue by examining the effect of the spin-orbit coupling onto those mechanisms.

As explained in Sec. 6.2, phonon-mediated transitions between the Rashba bands are suppressed due to energy-momentum conservation completely independently of the relative spin alignments. If we assume a high-purity sample and ignore carrier-impurity scattering for the moment, then inter-carrier scattering is the only way of inducing transitions between states of opposite chirality. We therefore have to check how the addition of spin-orbit coupling affects these transitions.

It is important to understand that the momentum-dependent spin mixing of the Rashba coupling is already taken into account explicitly in our calculations in Sec. 6.3 and App. B.3, where we include the spinor part of the wavefunctions in the matrix element. It will therefore only be necessary to probe the effect of momentum-dependent spin mixing to other Bloch bands and not between the two Rashba bands.

Phonon-mediated spin-orbit induced spin-flips can be described using the language of the Elliot-Yafet mechanism [13], where the momentum-dependent spin mixing to other Bloch bands is referred to as the Elliot contribution [49] and the effect of the phonon-modulated spin-orbit interaction referred to as the Overhauser contribution [105]. For the inter-carrier scattering, the Overhauser part is absent, and we will have to focus on the Elliot contribution only. We will now show that this vanishes up to first order in q. As we explain in Sec. 6.3, inter-carrier scattering is dominated by low-q transitions and therefore incompatible with this q dependence of the Elliot contribution.

We determine the scattering rate for spin-reversing inter-band processes by adding the spin mixing to another Bloch band to the transition-matrix element in Eq. (6.9). This is done by adding the product of the lattice-periodic part $u_{\mu\mathbf{k}}$ of the Bloch functions $\varphi_{\mu\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{kr})\, u_{\mu\mathbf{k}}(\mathbf{r})$ of initial and final state to the real-space integral in the expectation value. In momentum space, this results in multiplying the outcome of the result from Sec. 6.3 with the Fourier transform of the product of the initial and final $u_{\mu\mathbf{k}}$,

$$\int \mathrm{d}^2 r \exp\left(i\mathbf{qr}\right)\, u^*_{\mu\mathbf{k}} u_{\mu'\mathbf{k}+\mathbf{q}}\,. \tag{B.101}$$

Note that the labels μ and μ' refer to both the Bloch-band index and the pseudospin index that labels the spin degree of freedom. We can expand this integral around $q = 0$ and examine the leading-order terms. The zeroth order term is the overlap of the two wavefunctions with same momentum and opposite spin index and is zero as the $u_{\mu\mathbf{k}}$ are orthogonal. The term in first order integrates over the $\mathbf{r}$ vector,

$$i\mathbf{q}\int \mathrm{d}^2 r\, u^*_{\mu\mathbf{k}}\, \mathbf{r}\, u_{\mu'\mathbf{k}+\mathbf{q}}\,. \tag{B.102}$$

This term however must vanish because it can only exist in a system that breaks inversion symmetry[1]. Therefore, we can conclude that the lowest-order term in the small-q expansion is proportional to q^2.

This is the main reason why typically phonons are considered as the main source for spin-orbit induced spin-flips, as they are able to provide sufficient change of momentum to be compatible with the vanishing of the momentum-dependent spin mixing for small momentum transfer[2]. In other words, we can expect the spin mixing to have a similar effect in inhibiting inter-carrier transitions as the opposing spin alignments of the Rashba spin texture had in Sec. 6.3. This leaves us to conclude this section by summarising that spin-orbit coupling (Rashba coupling and to other Bloch bands) is ineffective in enabling inter-band transitions to opposite spin states.

[1] Note that we refer to the symmetry within the 2D plane, while obviously the Rashba interaction breaks the inversion symmetry in the direction perpendicular to the plane

[2] Check Ref. [13] for a detailed discussion of the low-q expansion of the Elliot-Yafet mechanism.